

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2004, 08:18:49 ; Search time 700.638 Seconds  
(without alignments)  
7148.662 Million cell updates/sec

Title: US-09-855-340A-1  
Perfect score: 1179  
Sequence: 1 grgtgatcgagaaaggg.....tggggcgcagcagcatga 1179

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : N\_Geneseq\_25Jan04:\*
- 1: Geneseqn1980s:\*
  - 2: Geneseqn1990s:\*
  - 3: Geneseqn2000s:\*
  - 4: Geneseqn2001as:\*
  - 5: Geneseqn2001bs:\*
  - 6: Geneseqn2002s:\*
  - 7: Geneseqn2003as:\*
  - 8: Geneseqn2003bs:\*
  - 9: Geneseqn2003cs:\*
  - 10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1179	100.0	1179	6	AAD25931
2	1179	100.0	4388	5	AAS08694
3	61.6	5.2	2404	2	AA337131
4	60.4	5.1	2000	7	ADA71938
5	56.6	4.8	1161	6	AD21926
6	56.6	4.8	1188	6	AD21912
7	55.8	4.7	27541	4	AD17185
8	55.8	4.7	125401	4	AD17186
9	55.2	4.7	114955	2	AA53491
10	55	4.7	114955	2	AA53491
11	53.2	4.5	2000	7	ADA71938
12	53	4.5	504	7	AD71938
13	53	4.5	82746	7	AD71938
14	52.8	4.5	24379	2	AA161224
15	52.8	4.5	24379	2	AA161224
16	52	4.4	88421	6	AA25925
17	51.4	4.4	516	2	AA47574
18	51.4	4.4	516	6	AA47574
19	51.4	4.4	516	6	AA47574
20	51.4	4.4	516	6	AA47574
21	51.4	4.4	516	6	AA47574
22	51.4	4.4	516	6	AA47574
23	51.4	4.4	582	6	AA47574

24	51.4	4.4	582	8	ADB78877
25	51.4	4.4	7065	8	AAF88577
26	51.4	4.4	7065	8	ADB78878
27	50.8	4.3	2136	7	ACA37643
28	50.6	4.3	1677	7	AA161210
29	50.6	4.3	82746	7	AA161224
30	50.4	4.3	11238	7	AA161224
31	50.4	4.3	47981	4	AAF30757
32	50.4	4.3	60196	7	AA161224
33	50	4.2	1291	4	ABL15825
34	50	4.2	3498	4	ABL15825
35	50	4.2	3790	4	ABL15824
36	49.8	4.2	4770	7	AA161172
37	49.8	4.2	109519	5	AA161172
38	49.6	4.2	9222	7	AA161171
39	49.6	4.2	135638	7	ABX34289
40	49	4.2	3197	6	ABK84161
41	49	4.2	14055	7	AA161170
42	48.6	4.1	1137	7	ABX56045
43	48.6	4.1	1189	7	ABT33357
44	48.6	4.1	1290	7	AA161172
45	48.6	4.1	53789	2	AAV21187

ALIGNMENTS

RESULT 1  
AAD25931  
ID AAD25931 standard; DNA; 1179 BP.  
XX AC AAD25931;  
XX DT 26-MAR-2002 (first entry)  
XX DE Micromonospora carbonacea pMLP1 integrase DNA.  
XX KW Integrase; int; excisionase; xis; integrase attachment site; attP; pMLP1;  
XX KW site-specific integration; hybrid antibiotic; metabolic product;  
XX KW secondary metabolic pathway; ds.  
XX OS Micromonospora carbonacea.  
XX PN WO200187936-A2.  
XX PD 22-NOV-2001.  
XX PF 15-MAY-2001; 2001WO-US015760.  
XX PR 17-MAY-2000; 2000US-0204670P.  
XX (SCHE ) SCHERING CORP.  
XX Hosted TV, Horan AC;  
XX WPI; 2002-082983/11.  
XX Novel polynucleotides encoding integrase, excisionase and an integrase attachment site isolated from a lysogenic phage pMLP1, useful for transforming an actinomycete.  
XX Claim 5; Page 32; 34pp; English.  
XX The present invention relates to novel polynucleotides encoding integrase (int) and excisionase (xis) and an integrase attachment site (attP) which are isolated from pMLP1, a bacteriophage (lysogenic phage) isolated from Micromonospora carbonacea var. africana. Polynucleotides of the invention are useful for transforming an actinomycete with a vector. They are also useful for creating vectors for site-specific integration into host chromosomes. The integrating vectors are used to express actinomycete genes, manipulate secondary metabolic pathways and create new metabolic products such as hybrid antibiotics. The present sequence is pMLP1 integrase DNA from Micromonospora carbonacea

```

XX SQ Sequence 1179 BP; 214 A; 370 C; 410 G; 185 T; 0 U; 0 Other;
Query Match 100.0%; Score 1179; DB 6; Length 1179;
Best Local Similarity 100.0%; Pred. No. 6.2e-207;
Matches 1179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGTGATCGAGAGAACGGCCCGTCTACCGATTCCGGACCTCGTTCGGTAAAG 60
DB 1 GTGTGATCGAGAGAACGGCCCGTCTACCGATTCCGGACCTCGTTCGGTAAAG 60
QY 61 GTACCATTCAGACCGGTATTCGACGAGAACCGCCCAAGAAATGCGATGTCAGTTTC 120
DB 61 GTACCATTCAGACCGGTATTCGACGAGAACCGCCCAAGAAATGCGATGTCAGTTTC 120
QY 121 GTGTGAGAGAGTTGTCAGGGCAACCGCTCATGCGCGCGCGGTGATGATTCCTCGCC 180
DB 121 GTGTGAGAGAGTTGTCAGGGCAACCGCTCATGCGCGCGCGGTGATGATTCCTCGCC 180
QY 181 GATTTCGTGGGGAGTGTGTGCGAGCTACGAAAGACGCTGAAACCGACCGCGGTGAAC 240
DB 181 GATTTCGTGGGGAGTGTGTGCGAGCTACGAAAGACGCTGAAACCGACCGCGGTGAAC 240
QY 241 TCGGAGGCAACCGGATCCGCAACCACTCTGCGCCATCTCGCCCATCTCACCCCTGAC 300
DB 241 TCGGAGGCAACCGGATCCGCAACCACTCTGCGCCATCTCGCCCATCTCACCCCTGAC 300
QY 301 GAGTGTGACGGCAGGTACCCAGACAGTGGGTCAACCTGTGAGCGCGCGTGGCCCG 360
DB 301 GAGTGTGACGGCAGGTACCCAGACAGTGGGTCAACCTGTGAGCGCGCGTGGCCCG 360
QY 361 TGGCGGAGTCCAGCGGGGTGCTGGAAGCGCTGGCAGGAGACGATCAGCAACTGC 420
DB 361 TGGCGGAGTCCAGCGGGGTGCTGGAAGCGCTGGCAGGAGACGATCAGCAACTGC 420
QY 421 CACGCGCTGTGCAACAGATCTGCGCGCGCGATCGCGGAAACCGATCAGGCTCAAC 480
DB 421 CACGCGCTGTGCAACAGATCTGCGCGCGCGATCGCGGAAACCGATCAGGCTCAAC 480
QY 481 CCGTGTCTTCGACGATGCTGCGCGCGCGAGCGGAGATGATGATTCCTGAGCGAC 540
DB 481 CCGTGTCTTCGACGATGCTGCGCGCGCGAGCGGAGATGATGATTCCTGAGCGAC 540
QY 541 CCGGAGATCGGTCTGCTTATCACGCGCTTCCGCGCGCTTCCGCGCGCTGCTGATGTC 600
DB 541 CCGGAGATCGGTCTGCTTATCACGCGCTTCCGCGCGCTTCCGCGCGCTGCTGATGTC 600
QY 601 CTGTGTGCGCGCGCTGAGTGGGTGAGCGGATCGCGCGCGCGCGCGGTGCGAC 660
DB 601 CTGTGTGCGCGCGCTGAGTGGGTGAGCGGATCGCGCGCGCGCGCGGTGCGAC 660
QY 661 CTGTGTGCGCGCGCGCGCTGAGTGGGTGAGCGGATCGCGCGCGCGCGGTGCGAC 720
DB 661 CTGTGTGCGCGCGCGCGCTGAGTGGGTGAGCGGATCGCGCGCGCGCGGTGCGAC 720
QY 721 GGAGAGTCTGCTTCCAGTTCGCGGAGACCGGAGCGCGCGCGCGCGGTGCGAC 780
DB 721 GGAGAGTCTGCTTCCAGTTCGCGGAGACCGGAGCGCGCGCGCGGTGCGAC 780
QY 781 ACAGAAATGCTCTACTGCTTACGCCACTCATCTCCGCGGAGTTCGGGTCAAG 840
DB 781 ACAGAAATGCTCTACTGCTTACGCCACTCATCTCCGCGGAGTTCGGGTCAAG 840
QY 841 TTACCGCGCGCGGAGTGGTAAAGACCGCAATTTCCGCGCGGTTCGGGTCAAG 900
DB 841 TTACCGCGCGCGGAGTGGTAAAGACCGCAATTTCCGCGCGGTTCGGGTCAAG 900
QY 901 CGGTGCGGAGAACCGCGGCTTCGGGCTTACGATTCAGCAATCTCGGCGACACTCACGG 960
DB 901 CGGTGCGGAGAACCGCGGCTTCGGGCTTACGATTCAGCAATCTCGGCGACACTCACGG 960
QY 961 GCGATCTGATTTCTGCGCGCGGTGCTGCGGATCTCCCGCGCGCTCGGTCACTCG 1020
DB 961 GCGATCTGATTTCTGCGCGCGGTGCTGCGGATCTCCCGCGCGCTCGGTCACTCG 1020

Db 961 GCGATCTGATTTCTGCGCGCGGTGCTGCGGATCTCCCGCGCGCTCGGTCACTCG 1020
QY 1021 TCGATCCGGTACGAGATCTGTTACGGGACCTGCTGAGGAGTGGACGAGGATC 1080
DB 1021 TCGATCCGGTACGAGATCTGTTACGGGACCTGCTGAGGAGTGGACGAGGATC 1080
QY 1081 CTGCGCGGATCGAGAGGCGATGCGCGCGGTGCGGCTGAGGACCTGAGGCGGAACTC 1140
DB 1081 CTGCGCGGATCGAGAGGCGATGCGCGCGGTGCGGCTGAGGACCTGAGGCGGAACTC 1140
QY 1141 GACGAGAGCTGACGAGCTGTTGGCGGACGAGCATGA 1179
DB 1141 GACGAGAGCTGACGAGCTGTTGGCGGACGAGCATGA 1179

RESULT 2
AAS08694
ID AAS08694 standard; DNA; 4388 BP.
XX AC AAS08694;
XX DT 11-SEP-2003 (revised)
XX DT 26-SEP-2001 (first entry)
XX XX Micromonospora DNA encoding integrase enzymes.
XX XX Everninomicin; antibiotic; bottle-neck gene; orthomycin; fermentation;
XX XX integrase; ds.
XX OS Micromonospora sp. ATCC 39149.
XX FH Key Location/Qualifiers
XX CDS 963..1387
XX FT /*tag= a
XX FT /product= "Integrase #1"
XX FT 1394..2572
XX FT /*tag= b
XX FT /product= "Integrase #2"
XX FT 2570..2799
XX FT /*tag= c
XX FT /note= "AttB/AttP region of integrase action"
XX FT 27114..2715
XX FT /*tag= d
XX FT /label= Insertion_juncture
XX FT /note= "Site of integrase activity"
XX PN WO200151639-A2.
XX PD 19-JUL-2001.
XX PF 12-JAN-2001; 2001WO-US001187.
XX PR 12-JAN-2000; 2000US-0175751P.
XX PA (SCHE ) SCHERING CORP.
XX PI Hosted TJ, Horan AC, Wang TX;
XX DR WPI; 2001-442147/47.
XX DR P-PSDB; AAU04900; AAU04912.
XX PT New nucleic acid molecules encoding everninomicin pathway gene products,
XX PT useful for improving yields of everninomicin, to produce new
XX PT everninomicin and as probes to identify homologous sequences.
XX PS Claim 26; Fig 7; 109pp; English.
XX CC The sequence encodes 2 integrases which permit site specific integration
XX CC of a vector into an actinomycete, especially a Micromonospora, genome.
XX CC The invention relates to nucleic acids and vectors comprising a M.
XX CC carbonacea everninomicin biosynthetic pathway resistance gene product
XX CC useful for selecting for a transfected or transformed host cell. An
XX CC integrative version of the vector is useful for introducing a

```

CC everninomicin pathway gene (a bottle-neck gene) into an actinomycete of  
CC the genus Micromonospora. The DNA encoding the biosynthetic proteins is  
CC useful for synthesising novel everninomicin-related compounds, arising  
CC from modifications of the DNA sequence designed to change glycosyl and  
CC modified orsellinic acid groups contained in everninomicin, for  
CC expressing functional or mutant everninomicin biosynthetic enzyme for  
CC evaluation, diagnosis and preferably biosynthesis of everninomicin or  
CC other secondary metabolic products, improving the yield of everninomicins  
CC and to produce novel everninomicins, and also as a hybridisation probe to  
CC identify homologous sequences. The encoded polypeptides are useful for  
CC combinatorial biosynthesis to generate libraries of orthomycins, e.g.  
CC everninomicin analogues/homologues and drug discovery. The DNA encoding  
CC the integrase allows for increasing a given gene dosage. The integrative  
CC vector can be used to permanently integrate copies of a heterologous gene  
CC of choice into chromosomes of different hosts and to integrate genes  
CC which increase the yield of known products or to generate novel products  
CC such as hybrid antibiotics or other novel secondary metabolites. The  
CC vector can also be used to integrate antibiotic resistance genes in order  
CC to carry out bioconversions with compounds to which the strain is  
CC normally sensitive and is thus useful in fermentation processes involving  
CC e.g. Streptomyces antibioticus. (Updated on 11-SEP-2003 to standardise OS  
CC field)  
XX  
SQ Sequence 4388 BP; 714 A; 1468 C; 1506 G; 694 T; 0 U; 6 Other;

Query Match 100.0%; Score 1179; DB 5; Length 4388;  
Best Local Similarity 100.0%; Pred. No. 6.1e-207;  
Matches 1179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGTGATCGAGAAAGCGGGCCGCTACCGCATTCGGGACCTCGTTCGGGGTAAAG 60  
DB 1394 GTGTGATCGAGAAAGCGGGCCGCTACCGCATTCGGGACCTCGTTCGGGGTAAAG 1453  
QY 61 GTCCACATTCAGACCGGTTATCCGACGAAGACAGCGGCAAGATGCGATGTCAGTTC 120  
DB 1454 GTCCACATTCAGACCGGTTATCCGACGAAGACAGCGGCAAGATGCGATGTCAGTTC 1513  
QY 121 CGTGGGAGGAGTTCAGGGCAACCGCTCATCGCGCGGGGTCAGATTACCTCGCC 180  
DB 1514 CGTGGGAGGAGTTCAGGGCAACCGCTCATCGCGCGGGGTCAGATTACCTCGCC 1573  
QY 181 GATTTCGTGGGGAGTGTGGCGGAGTACGAAAGAGCGCTGAACCGACCGCGTGAAC 240  
DB 1574 GATTTCGTGGGGAGTGTGGCGGAGTACGAAAGAGCGCTGAACCGACCGCGTGAAC 1633  
QY 241 TCGGAGGGCAACCGGATCCGAACCACTCTGTGCCATATCTGGCCATCTCACCTTTGAC 300  
DB 1634 TCGGAGGGCAACCGGATCCGAACCACTCTGTGCCATATCTGGCCATCTCACCTTTGAC 1693  
QY 301 GAGCTGGACGGGACGTTCAACCGAGTGGGTCAACGACCTGGAGCGCGGTCGGCCCG 360  
DB 1694 GAGCTGGACGGGACGTTCAACCGAGTGGGTCAACGACCTGGAGCGCGGTCGGCCCG 1753  
QY 361 TGGCGGAGTCCACCGGGGTCGTGGGAAGCGCTGGCAGCGAAGAGATCAAGATTCCTGAGCGAC 420  
DB 1754 TGGCGGAGTCCACCGGGGTCGTGGGAAGCGCTGGCAGCGAAGAGATCAAGATTCCTGAGCGAC 1813  
QY 421 CACGCGCTGCTGCACACGATCTGGGCGCGGGATCGCGGCAACCGATCAGGCTCAAC 480  
DB 1814 CACGCGCTGCTGCACACGATCTGGGCGCGGGATCGCGGCAACCGATCAGGCTCAAC 1873  
QY 481 CCGTCTCTTCGACGATCTCTCCCGCGCGGAGCGGAAAGAGATCAAGATTCCTGAGCGAC 540  
DB 1874 CCGTCTCTTCGACGATCTCTCCCGCGCGGAGCGGAAAGAGATCAAGATTCCTGAGCGAC 1933  
QY 541 CCGGAGATCGGTGGCTTATCAGCGGCTTCGCGGCACTGGCGACCGCTCGTCACTG 600  
DB 1934 CCGGAGATCGGTGGCTTATCAGCGGCTTCGCGGCACTGGCGACCGCTCGTCACTG 1993  
QY 601 CTGTGGCGACCGGTCGTGAGTGGGGTGAAGCGGATCGGCTCGCGCGGCGGGTCGAC 660  
DB 1994 CTGTGGCGACCGGTCGTGAGTGGGGTGAAGCGGATCGGCTCGCGCGGCGGGTCGAC 2053

QY 661 CTGCTCGCGCGCGGCTCCCGCTGACCGTCTGTCAGCAGCTCCAGAGGCTGGCCAGGACG 720  
DB 2054 CTGCTCGCGCGCGGCTCCCGCTGACCGTCTGTCAGCAGCTCCAGAGGCTGGCCAGGACG 2113  
QY 721 CGAGAGCTCGTCTTCCAGTCCCGAAGACCGAAGGCGCGGCGCACGGTCAGTTTCACC 780  
DB 2114 CGAGAGCTCGTCTTCCAGTCCCGAAGACCGAAGGCGCGGCGCACGGTCAGTTTCACC 2173  
QY 781 ACGAAGTCTGCTTACTGCTTACCCACTCATCCCGAAGAAAAGTGAAGGTCGTCG 840  
DB 2174 ACGAAGTCTGCTTACTGCTTACCCACTCATCCCGAAGAAAAGTGAAGGTCGTCG 2233  
QY 841 TTCACCGCGCGAAGCGGATGTAAGGACGCGCAATTCGCGCGGATCTGGTCAAG 900  
DB 2234 TTCACCGCGCGAAGCGGATGTAAGGACGCGCAATTCGCGCGGATCTGGTCAAG 2293  
QY 901 GCGTGCAGGAGACCGGGCTTCGCGGCTTACGCAATTCACGATCTGCGGCGACACTCAGCGG 960  
DB 2294 GCGTGCAGGAGACCGGGCTTCGCGGCTTACGCAATTCACGATCTGCGGCGACACTCAGCGG 2353  
QY 961 GCGATCTGATTTCTGCGGGGCTCCGCTGTCGGCGATCTCCCGCGCTCGGTCACTCG 1020  
DB 2354 GCGATCTGATTTCTGCGGGGCTCCGCTGTCGGCGATCTCCCGCGCTCGGTCACTCG 2413  
QY 1021 TCGATCGCGTTCAGGATCTGCTGTACGGGCACTTCGCTGAGGAGTCCGACGAGGGATC 1080  
DB 2414 TCGATCGCGTTCAGGATCTGCTGTACGGGCACTTCGCTGAGGAGTCCGACGAGGGATC 2473  
QY 1081 CTGCGCGGATCGAGGAGGATGCGCGGCTCGGGCTCGGGCTGAGGACCTGAGGCGGAATC 1140  
DB 2474 CTGCGCGGATCGAGGAGGATGCGCGGCTCGGGCTCGGGCTGAGGACCTGAGGCGGAATC 2533  
QY 1141 GACGAGGAGTGCAGGACGCTGTTGGCGACGCGACATGA 1179  
DB 2534 GACGAGGAGTGCAGGACGCTGTTGGCGACGCGACATGA 2572  
RESULT 3  
AAAX37131  
ID AAAX37131 standard; DNA; 2404 BP.  
XX AC AAAX37131;  
XX DT 05-JUL-1999 (first entry)  
XX DE Mycobacteriophage M56 DNA attP-int region.  
XX DE DNA integration; Mycobacterium; bacteriophage; phage attachment site;  
XX KW attP; promoter; integrase; recombinant; transformation efficiency;  
XX KW vaccine; ss.  
XX OS Mycobacterium sp.  
XX PN WO9907861-A1.  
XX PD 18-FEB-1999.  
XX PF 06-AUG-1997; 97WO-PT000005.  
XX PR 06-AUG-1997; 97WO-PT000005.  
XX (MEDI-) LAB MEDINFAR-PROD FARMACEUTIC LD7A.  
XX PI Frazao Moniz Pereira JA, Freitas Vieira A;  
XX PI Ribeiro Dos Santos Anes EM, Da Costa Garcia MA, Da Silva Alves PJ;  
XX WPI; 1999-180493/15.  
XX DR P-PSDB; AAY09007.  
XX PT A new system for integrating DNA into mycobacterium species - allows the  
XX PT stable construction of a vaccine vehicle for long-term antigen gene  
XX PT expression.

PS Example 1; Fig 3; Sipp; English.

XX The invention relates to the integration of a DNA fragment into a  
CC specific site of the Mycobacterium genome, using the integrating DNA  
CC functions of a bacteriophage. A genetic system for integrating the DNA  
CC comprises: (a) DNA containing IF of a bacteriophage linked to the DNA to  
CC be expressed under control of a promoter; or (b) an integrative plasmid  
CC carrying the phage attachment site (attP) and the DNA to be expressed  
CC under control of a promoter, and a helper plasmid encoding an integrase.  
CC The system can be adapted for other bacteria such as *E. coli*, *Salmonella*  
CC spp., *Vibrio* spp., *Shigella* spp., *Listeria* spp., *Streptococcus* spp.,  
CC *Lactobacillus* spp., *Corynebacterium* spp., and *Streptomyces* spp. The  
CC recombinant mycobacterium is used as a vaccine. Transformation efficiency  
CC using this integration system is higher than that of prior art DNA  
CC integration using double homologous recombination

XX Sequence 2404 BP; 437 A; 714 C; 791 G; 462 T; 0 U; 0 Other;

Query Match 5.2%; Score 61.6; DB 2; Length 2404;  
Best Local Similarity 48.5%; Pred. No. 0.022;  
Matches 391; Conservative 0; Mismatches 364; Indels 51; Gaps 6;

QY 257 TCGGCAACCACTCTGCTCCCTACTCGCCATCTCACCCTTGACAGCTGGAGCGGACG 316  
DB |||||  
QY 995 TCGGCAAGCATCGAACCACTCGCCGACATCCCTTGTGGAAGCTTCGGAAGAG 1054  
DB |||||  
QY 317 TCACCCAGCAGTGGTCAACAGCTGGAGCGCGCTGCGCGGAGTCCACGC 376  
DB |||||  
QY 1055 ACATCGCCGCTGGTGAAGTCTATGAACCAACCGCGGCCG-----CGAGCGCAACG 1109  
QY 377 GGGTGTGCGAAGCGCTGGACGGAAGCATGACCACTGCCAGCGCTGCTGCACA 436  
DB |||||  
QY 1110 GCACGCCCGCAACCTCCGCAACAAATACGG---GTTCTATCGGGGGCACTGAACGC 1166  
QY 437 CGATCTGGCGCGCGCATCGCGCGAAACGGATCAGCTCAACCGTGTCTTCGACA 496  
DB |||||  
QY 1167 CGCGCTCCCGATCTTTCACCACTCGGTGGCGCGCGCTGCGCGGTGGAA 1226  
QY 497 TGCTGCCCGCGCGGACGCGGAAGATGAAGTCTTGACGACCGCGAGATCGGTGCG 556  
DB |||||  
QY 1227 CGCTG-----AGGACGACGACGAGATCGCGATCGCTCAACCGCGGAGTTCGACCGCG 1279  
QY 557 TTATCAGCGGCTTCGCGCGCACTGGCGACCGCTGCTGCTGCTGGGACCGGTC 616  
DB |||||  
QY 1280 TCCGCGACCGGTGACACCTCACTGGAAGTGTGTTCAATGTTGTCGACCGGT 1339  
QY 617 TGAGTGGGTGAGCGGATCGGCTCGCGCGCGCGGTCGACCTGCTCGCGCGCGCG 676  
DB |||||  
QY 1340 TCGGTGGGTGAGGTATCGCGCTGCGCGCGCGGATGTTGGAGATGTCACCA 1399  
QY 677 CCGGCTGACCGTCTGACGAGCTCCAGGAGTGGCGACGCGGAGAGCTGCTTCC 736  
DB |||||  
QY 1400 TCAGG-----GTGCGGACGCGTGGAGTACTGCTCGCGCGGTATGTTGG 1447  
QY 737 AGTCCCGAAGACCGCGAAGCGCGGACCGGTCTGCTGCTGCTGCTGCTGCT 796  
DB |||||  
QY 1448 GCGCGCGCGAAGCAACCGTTCGCGCGCGCGGTGATGCTGCGCGCGGATGTTGGAGC 1507  
QY 797 TGCTTACCGCACTCATCGCGGAAAGAAAGTCAAGGTCGTTTCAACCGCGCGAAG 856  
DB |||||  
QY 1508 GGCTG-----GACTTGTGCAACGAGTTGTTTCTGTCATACCGATG 1549  
QY 857 GCGGATGTTAAGACCGCAATTCGCGCGGATCTGGGTCAAGGCTGCG-----AGG 910  
DB |||||  
QY 1550 GTGACCGCGTCAAGTATCCCGGCTTCTGCGTGGGTGTTGGAATCCGCTGCGGAGAG 1609  
QY 911 AAGCGGCTTCCGGCTTACGATTCACCATCTGCGGACACTCACCGCGGATCTCTGA 970  
DB |||||  
QY 1610 CTGCTGTGTTCCGCGGCTTCTCCGACGATCTGCGGACACGATGCTGTCGCGAGC 1669  
QY 971 TTTTTCGCGGCGTTCGCTGTGCGGATCTCCCGCGCGCTCGGTCTACTGTCGTCGCGG 1030  
DB |||||  
QY 1670 TAACGGCGGACACCGCGTACGATTTGTCTCGCGAGCTGGGTCTATGATGCTGCA 1729

QY 1031 TCACGATCTGCTGTACGGCACCTG 1056  
DB |||||  
DB 1730 TCACGGTGGACAGCTACACGATGTG 1755

RESULT 4

ADA71938/c

ID ADA71938 standard; DNA; 2000 BP.

XX ADA71938;

XX ADA71938;

XX 20-NOV-2003 (first entry)

XX Rice gene, SEQ ID 5263.

XX Plant; bacterial infection; fungal infection; viral infection; rice;

XX gene; ds.

XX Oryza sativa.

XX WO2003000898-A1.

XX 03-JAN-2003.

XX 22-JUN-2001; 2001WO-IB001105.

XX 22-JUN-2001; 2001WO-IB001105.

XX (SYGN) SYNGENTA PARTICIPATIONS AG.

XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;

XX Katagiri F, Qian S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;

XX WPI; 2003-175290/17.

XX Identifying at least one gene involved in plant resistance or response to  
XX pathogenic infection for conferring resistance or tolerance to a plant to  
XX bacterial, fungal or viral infection by determining or detecting plant  
XX gene expression.

PS Claim 27; SEQ ID NO 5263; 899pp; English.

XX The present invention relates to a method (M1) for identifying genes  
XX involved in plant resistance or response to pathogenic infection. M1  
XX comprises identifying a gene whose expression is significantly altered in  
XX the incompatible interaction of plant gene expression relative to  
XX expression of the gene in an uninfected plant, in a mutant plant that  
XX does not express a gene associated with response to pathogenic infection,  
XX or in a corresponding incompatible or compatible interaction. (M1) is  
XX useful for conferring resistance to resistance or tolerance to a plant to  
XX bacterial, fungal or viral infection. The present sequence was used to  
XX illustrate the invention.

XX Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match

Best Local Similarity 5.1%; Score 60.4; DB 7; Length 2000;

Matches 65; Conservative 277; Mismatches 263; Indels 3; Gaps 1;

QY 204 GAGCTACGAAAGACCTGAAACCGACCGCTGAACTCGGAGGCAACCGATCCGCA 263  
DB |||||  
QY 653 GAGSMYKYSKSAKCKRTRTMTSYMTGMYTSSYKSMWTSKMSYKMTCTYTT 594  
QY 264 CCACCTCTCTCCCATCTCGGCACTCTCCCTTACGAGCTGGAGCGGACAGTCA 323  
DB |||||  
QY 593 SMKGSTRSRKMGWSGMSRMWRWKKRMYKMKCTWRRCMCRVGYTYTTTSR 534  
QY 324 GCAGTGGTCAAC---GACTGAGCGCGGCTCGCGCGTGGCGGAGTCCACGCGG 380  
DB |||||  
QY 533 SRMYTGGRYARTSKRRYRWYKRYWYKRYWYKRYWYKRYWYKRYWYKRYWYK 474  
QY 381 TCSTCGAAAGCGCTCGGACGGAAGACGATACGAACTGCCACGCGCTCTGCA 440

```

Db 473 KAAYSGMMYWRKYKSWMRMSTYXWMSMYKCRSNKYGAKGYGCKWTCYSGYMKW 414
Qy 441 CTGGCGGCGGCGATCGCGGGAACGATCAGCTCAACCGCTGCTCTCGAGATGCT 500
Db 413 YTYGSKYKRCYKMYKMYKGMWYMYYSAYSSMMTWYYYAKYKWKYKRGTMWY 354
Qy 501 GCCCGGCGGAGCGGAAAGATGAAGTTCCTGAGCGACCCGAGATCGGCTTAT 560
Db 353 GKSYKKYCTWVCYMKCMRCYWRKQMKRKTYSKRCYOWRYATCYWCCYKRGWYSR 294
Qy 561 CAGGCGCTTCCCGGCACTGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 620
Db 293 SMRTAGKWRKWSRWCBSYKMYKMKWKKSYNYSYNSYNSYNSYNSYNSYNSY 234
Qy 621 GTGGGTGAGCGATCGGCTCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTG 680
Db 233 STSRRAKWRACRYSACRYSRTSYCGSYGSSKWKYMSKSMRWTCSWSGCCYT 174
Qy 681 GCTGACGCTGCTGAGAGCTCAGGAGCTGCGGAGAGAGAGAGAGAGAGAGAG 740
Db 173 CYTGAMCWSGSMYMGSCYTRGWKWSKYSMCKKYCSCCTKYCSYTGVRKYKY 114
Qy 741 GCCGAAGACCGGAGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTG 800
Db 113 KYSYKCYCYCYWYNSYMYKMKCRSCSWSMSCAVCSTSTSRMWSMYAAMGMC 54
Qy 801 TAGGCCAC 808
Db 53 GSSGMVYM 46

```

## RESULT 5

```

AAD21926
ID AAD21926 standard; DNA; 1161 BP.
XX
AC AAD21926;
XX
DT 12-FEB-2002 (first entry)
XX
DE M. rosaria pMR2 plasmid site-specific integrase gene, int.
XX
KW Plasmid pMR2; vector construction; secondary metabolic pathway;
XX antibiotic; int gene; site-specific integrase; ds.
XX
OS Micromonospora rosaria.
XX
FN WO200175116-A2.
XX
PD 11-OCT-2001.
XX
PF 29-MAR-2001; 2001WO-US010321.
XX
PR 04-APR-2000; 2000US-0194461P.
XX
PA (SCHE ) SCHERING CORP.
XX
FI Hosted TJ, Horan AC;
XX
WPI; 2002-010791/01.
XX

```

```

XX New genes isolated from Micromonospora rosaria plasmid pMR2, useful in
XX constructing vectors for studying and expressing genes, or in
XX manipulating secondary metabolic pathways in actinomycetes.
XX
PS Claim 9; Page 33-34; 34pp; English.
XX
CC The invention relates to new genes isolated from Micromonospora rosaria
XX plasmid pMR2, and proteins encoded by such genes. The isolated genes of
XX the invention are useful in the construction of vectors, which can be
XX used in the study and expression of genes, in manipulating secondary
XX metabolic pathways in actinomycetes and in creating new metabolic
XX products such as hybrid antibiotics. The isolated genes can also be used

```

```

CC to create replicating, Escherichia coli-actinomycete shuttle, integrating
CC and intermycelial and intramycelial conjugation vectors for use in
CC actinomycetes. The present sequence is M. rosaria pMR2 plasmid site
CC specific integrase gene, int. This sequence is involved in pMR2 plasmid
CC integration
XX
SQ Sequence 1161 BP; 188 A; 398 C; 387 G; 188 T; 0 U; 0 Other;

```

```

Query Match 4.8%; Score 56.6; DB 6; Length 1161;
Best Local Similarity 46.2%; Pred. No. 0.19;
Matches 262; Conservative 0; Mismatches 299; Indels 6; Gaps 2;

```

```

Qy 545 AGATCGGTCGGTATACGCGGCTTCGCGCGCATCGCGGACCGCTCGCTGCTGCTG 604
Db 557 AGTCTGTCGCTGTCGAGGCCATCAACCCGCGCTACCGACGCGCTGCTGATCGCG 616
Qy 605 TGGGACCGGTCTGAGGTGGGTGAGCGATCGCGCTGCGCGCGCGCGCTGACCTGC 664
Db 617 CTGTGCGGCTACGCGCGGAGAGATCGCAGGACTCCGGGTGCAGACGTGACCTG- 675
Qy 665 TCGCGCGCGCGCGCTGACCGTCTGTCGAGCAGCTCCAGGAGTCCAGGAGTCCAGCAGGAG 724
Db 676 --ACCGAACACCATCACCGTCCGCAAGCCAGGTCGAACCGCTGCACGACAGGGGA 733
Qy 725 AGCTCGTCTTCCAGTTCGCGGAGACCGGAGGCGCGCGCATCGCGCTGCGCGCGCGCTGACCTGCA 784
Db 734 AGGCGTTTGACAAGATCCCAAGTCCGAGGCGGCAAGCAACCATGCCATCCCTCCC 793
Qy 785 AAGTCGCTCTACTCTTACGCCACTCATTCGCGGAGAAAGAAAGTACGAGGTCTGTTCA 844
Db 794 ACCTGTAACCGCTGATCCGCTGCACCTCGACGAGTTCCGCGGAGGATCGCTGTTG 853
Qy 845 CCGCGCGGAAAGCGGATGTTAGGACGCGCAATTTCCGCGGATCTGGTCAAGCGT 904
Db 854 TCAGCCGCGAGGATCACCTTCGCGGCGACACCTGTACGAGCGCTTGTAGCGGCTC 913
Qy 905 GCGAGGAGCGGCTTCGCGGTTACGCAATTCAGATCTCGGCACTCACTCAGCGGCGA 964
Db 914 GGGGAAAGTTCGGAATCGACACCTCACCTTCACGACCTGCGGCAACCGGTGACACCC 973
Qy 965 TCCTGATTTCTGCGGCGTCCGCTGTCGCGGATCTCCCGCGCTCGGTCTGCTGCTG 1024
Db 974 TCGCGCGGAGACCGGCGGACCTTGGCGGACCTGATGAAGCGGCTCGGCACTCGTCCA 1033
Qy 1025 TCGCGGTACGAGTCTGCTGACGCGCACCTGCTGAGGAGTCCGAGGAGGATCCTCG 1084
Db 1034 TGGCTG--CGGCTCGCGGTACTCCACGCGCTTGAAGTCTGACGCTGCTGACGCGGAGATGCCA 1090
Qy 1085 CGGCGATCGAGGAGGATGCGCGGCG 1111
Db 1091 AAGCCCTTTCGAGGCTGGCGGCGCAG 1117

```

## RESULT 6

```

AAD21912
ID AAD21912 standard; DNA; 11188 BP.
XX
AC AAD21912;
XX

```

```

XX 12-FEB-2002 (first entry)
XX
DE Micromonospora rosaria pMR2 plasmid DNA.
XX
KW Plasmid pMR2; vector construction; secondary metabolic pathway;
XX antibiotic; ds.
XX
OS Micromonospora rosaria.
XX

```

```

XX Key Location/Qualifiers
XX misc_feature 59..102
XX /tag= a
XX /note= "pMR2 plasmid attachment site encoding gene, att
XX p"

```



```

OS Streptomyces noursei.
XX Key Location/Qualifiers
PH CDS complement(454..1191)
FT /*tag= a
FT /product= "NysF protein"
FT /note= "CDS does not include start codon"
CDS complement(1275..3092)
FT /*tag= b
FT /product= "NysG protein"
FT complement(3070..4824)
FT /*tag= c
FT /note= "CDS does not include start codon"
FT /product= "NysH protein"
CDS 5122..6156
FT /*tag= d
FT /product= "NysD3 protein"
FT complement(6307..27541)
FT /*tag= e
FT /product= "NysI partial protein"
FT /note= "CDS does not include stop codon"
XX
XX W0200159126-A2.
XX
XX 16-AUG-2001.
XX
XX 08-FEB-2001; 2001WO-GB000509.
XX
XX 08-FEB-2000; 2000GB-00002840.
XX
XX 10-APR-2000; 2000GB-00009786.
XX
XX 14-APR-2000; 2000GB-00009387.
XX
XX (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
XX (SNTF) SINTEF STIFTELSEN IND TEK FORSK.
XX (ALPH-) ALPHARMA AS.
XX (SINV-) SINVENT AS.
XX (DZIE-) DZIEGLEWSKA H.
XX (ZORC-) ZOTCHEV S B.
XX (SEKU-) SEKUROVA O N.
XX (FJAE-) FJAEV E.
XX (BRAU-) BRAUTASET T.
XX (STRO-) STROM A R.
XX (VALL-) VALLA S.
XX
XX Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
XX Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
XX WPI; 2001-557614/62.
XX P-PSDB; AAE10138, AAE10139, AAE10140, AAE10141, AAE10142.
XX
XX New nystatin polyketide synthase polynucleotides and polypeptides, useful
XX as antibiotics and antifungals.
XX
XX Claim 2; Page 151-166; 266pp; English.
XX
XX The present invention relates to the cloning and sequencing of the gene
XX cluster encoding a modular type I polyketide synthase (PKS) enzyme
XX involved in the biosynthesis of the macrocyclic antibiotic nystatin. The
XX nystatin PKS is useful as antifungal antibiotics. The present sequence is
XX a Streptomyces noursei nys2 DNA of nystatin PKS gene cluster
XX
XX Sequence 27541 BP; 3517 A; 10766 C; 9529 G; 3729 T; 0 U; 0 Other;
XX
XX Query Match
XX Best Local Similarity 4.7%; Score 55.8; DB 4; Length 27541;
XX Matches 346; Conservative 0; Mismatches 417; Indels 10; Gaps 3;
XX
XX 297 TGACGAGCTGACGGGAGGTGACCCAGCTGGGTCAACGCTGAGGCGGCGCGG 356
XX 4651 TGTGTGATGGGTGGGGCTGGTGCAGATCGCGCGCGCTCGTGGGCGCGCT 4592
XX 357 CCGTGGCGGAGTCCACGGCGGTGCTGCGAAGCCGCTGCGAGCAAGATACAGAA 416
XX

```

```

Db 4591 GCGCCGCCGACCGGACCGGCGATGGGCGCGGATCTGCGTCCGCGCTCTTCCGCGCAT 4532
QY 417 CTGCGACGCGCTGTGCACACGATCTGCGGCGCGGCGATCGCGGCGAAACGATCAGGCT 476
Db 4531 CCGTGCATCTCCGCGCGCGGATCGCGGCTTCCGACCCCTCCCTGCTGACCCGCTC 4472
QY 477 CAACCCGCTGCTTTCGACGATGCTGCCCGGCGGAGCGGAAAGAGATGAAGATTCCTGAG 536
Db 4471 GGTGAACGACGTGCAGCAGGTGCAGAACCTCGCCAGACCGGCTTCGCGCATCGTCTG 4412
QY 537 CGACCCGCGAGATCGTGGCTTATCACGGCGCTTCCGCGCGACTGG---CGACCGCTCGT 593
Db 4411 CGCGCGCTGATGCTGCTGGGCGGCTGCTGCGCGCTGCGCGCAGGCTTCGCGCTGGC 4352
QY 594 CATGCTGCTGTTGCGGACCGGCTCTGAGGTGGGGTGAGCGCATCGGCGCTGCGCGCGGCG 653
Db 4351 ACTGCTGCTGGTGGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 4292
QY 654 GGTGACCTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 707
Db 4291 GATGGGACCGCTGTACCGCGCGGATGCAGCTGACCGCTGACCGCTGCGCGCGCGCTGCG 4232
QY 708 GCTGCGCAGCAGCGGAGAGCTCGCTTCCAGTCCCGGAGACCGCGGAGCGCGCGCGCG 767
Db 4231 CGAGCGGATCACCGGGTGGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 4172
QY 768 GGTGAGTTTACAGAGAAAGTGGTCTTACTGCTTACGCCACTCATCCCGGAAAGAAAAG 827
Db 4171 GCGTTTCGCGCAGACCAACACGCGTTCCTCGTGGTGGTGGTGGTGGTGGTGGTGG 4112
QY 828 TGACGAGTGTGTTTACCGCGCGGCGGAAAGCGGAGTGAAGGACGCGCAATTTCCGCGG 887
Db 4111 CGCGACGATGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4052
QY 888 GATCGGCTCAAGCGTGCAGGAGCGCGGCTTCCGCGCTTACGCAATTCACGATCTGCG 947
Db 4051 GACCGGCTCCACCGGATCGACGCGCGCGGATGCGCGATCGGCTCGCTCAGCGCGCTCT 3992
QY 948 GCACACTCAGCGCGGATCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1006
Db 3991 GAGCTATCTCTGCTGATCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3932
QY 1007 GCTCGGCTCACTGCTGATCGCGCTCACGGATCTGCTGTACGGGCACTCGCT 1059
Db 3931 GCGCGCGCGCGGCTGCTGCGCGCGGATCGCGGAGTCTCGACACCGGCT 3879
XX
XX AAD17186/c
XX ID AAD17186 standard; DNA; 125401 BP.
XX AC AAD17186;
XX
XX 29-NOV-2001 (first entry)
XX
XX Streptomyces noursei nystatin PKS gene cluster DNA.
XX
XX Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
XX antifungal; antibiotic; ds.
XX
XX Streptomyces noursei.
XX
XX Key Location/Qualifiers
XX CDS 6337..34771
XX /*tag= a
XX /product= "NysI complete protein"
XX 34792..51099
XX /*tag= b
XX /product= "NysJ protein"
XX 51155..57355
XX /*tag= c
XX /product= "NysK protein"
XX 57503..58687
XX

```

```

FT      /tag= d
FT      /product= "NysL protein"
FT      complement(58786..58980)
FT      /tag= e
FT      /product= "NysM protein"
FT      /note= "CDS does not include start codon"
FT      complement(59045..60241)
FT      /tag= f
FT      /product= "NysN protein"
FT      /note= "CDS does not include start codon"
FT      complement(60238..61296)
FT      /tag= g
FT      /product= "NysD2 complete protein"
FT      120628..121308
FT      /tag= h
FT      /product= "NysR4 (long) protein"
PN      WO200159126-A2.
XX      16-AUG-2001.
XX      08-FEB-2001; 2001WO-GB000503.
XX      08-FEB-2000; 2000GB-00002840.
XX      10-APR-2000; 2000GB-00008786.
XX      14-APR-2000; 2000GB-00009397.
XX      (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
XX      (SNTF) SINTEF STIFTTELSEN IND TEK FORSK.
XX      (ALPH-) ALPHARMA AS.
XX      (SINV-) SINVENT AS.
XX      (DZIE/) DZIEGLEWSKA H.
XX      (ZOTC/) ZOTCHEV S B.
XX      (SEKU/) SEKUROVA O N.
XX      (FJAE/) FJAEVIRIK E.
XX      (BRAU/) BRAUTASET T.
XX      (STRO/) STROM A R.
XX      (VALL/) VALLA S.
XX      Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
XX      PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX      WPI: 2001-557614/52.
XX      P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
XX      AAE10149, AAE10150.
XX      New nystatin polyketide synthase polynucleotides and polypeptides, useful
XX      as antibiotics and antifungals.
XX      Claim 1; Page 188-254; 266pp; English.
XX      The present invention relates to the cloning and sequencing of the gene
XX      cluster encoding a modular type I polyketide synthase (PKS) enzyme
XX      involved in the biosynthesis of the macrocyclic antibiotic nystatin. The
XX      nystatin PKS is useful as antifungal antibiotics. The present sequence is
XX      a Streptomyces noursei nystatin PKS gene cluster DNA
XX      Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 U; 0 Other;
XX      Query Match 4.7%; Score 55.8; DB 4; Length 125401;
XX      Best Local Similarity 44.8%; Pred. No. 0.24;
XX      Matches 346; Conservative 0; Mismatches 417; Indels 10; Gaps 3;
XX      QY 297 TCACGAGCTGGACGGGCGAGTTCACCCAGCAGTGGGTCAACGACCTGGAGCGCGCGCTCGG 356
XX      4651 TCTGTGATGGCGTGGTGGCGTGTGTCAGATCGCGCGCGCTCGGTGGCGCGCT 4592
XX      QY 357 CCCGTGGCGGAGTTCACGCGGGTCTGTGGAACCGCTGGCAGCAACGATCAGCAA 416
XX      DB 4591 GCGCGCGCGGACCGCAGCGGATGGCGCGGATCTGCGGTCCGCGCTTTCGCGCGCAT 4532
XX      QY 417 CTGCCACGCGCTGTGTCACACGATCTCGCGCGGCGGATCGCGCGGAAACGAGTCAAGGT 476

```

```

DB 4531 CCTGGACTTCTCCGCGCGGAGATCGGCGGTTCCGCAACCCCTCCCTGCTGACCCGGTCTC 4472
QY 477 CAACCCGTGCTCTTCGACGATCTGCCCGCGGCGGAGAGAGATGAAGTTCCTGAG 536
DB 4471 GGTGAACGAGTGCAGAGGTGCAGAACTTCGCCAGACCGGCTTCGGATCGTCTGTG 4412
QY 537 CGACCCGGAGATCGGTCGGCTTATCACGGCGGTTCCGCGCACTGG---CGACCGCTCGT 593
DB 4411 CGGCGCGCTGATGTGCTGGGACGCTGCTGTGGCGTGTGCGCAGGACGTTCCGCTGGC 4352
QY 594 CATGCTGCTGGTGGGACCGTCTGAGGTGGGTGAGCGATCGGCTCGCGCGCGCG 653
DB 4351 ACTGCTGCTGGTGGCGCTGTGCTGTGTGGCGGTCTGTTCCGCGCTGCTGTGTCGCCCG 4292
QY 654 GGTGACCTGCTCGCGCGCGCGCGGCTGACCGTCTGTC-----GAGCAGCTCCAGGA 707
DB 4291 GATGGGCACTGTACGCGCGGATGAGCTGACCTTGACCGGCTGGGCGGCTGCTGCG 4232
QY 708 GCTGGCCAGCAGGAGAGTCTGCTTCAGTCCGCGAGACCGGAGGGCGGCGCAC 767
DB 4231 CGAGGCGATCACCGGGGTGCGGTGCTCGGTCTTGTGTCGCGGACGACCGAGCGCGC 4172
QY 768 GGTGAGTTTCAACCAAGTTCGCTCTACTGCTTACGCCACTCATCGCGGAAAGAAAG 827
DB 4171 GCGTTTCGCGCAGACCAACGACGCGTCTCTGCTGTGTGTCGCGCGGTGGTTCGCTGAT 4112
QY 828 TGACGAGTCTGTTTACCGCGCGCGGAGGCGGATGTTAGGACGCGCAATTTCCGCGC 887
DB 4111 CGCGACGATGCTCCGCTGCTGCTGTGTGATGAACGTTTACCGGAGCCCTGCTGTG 4052
QY 888 GATCTGGTCAAGGCGTGCAGAGAACCGCGGCTTCCGGCTTACGATTCAGATCTCGC 947
DB 4051 GACCGCTCCACCGGATGACGCGCGCGGATGCGATCGCTCGCTCAGCGCGCTCCT 3992
QY 948 GCACATCAGCGCGGATGCTGATTTCTGCGG-GGCTCCGCTGTCGGGATCTCCGCC 1006
DB 3991 GAGCTATCTCGCTGATCTCTGATGTCGCTGATGTCGCTGCTGCTGCTGCTGCTGCT 3932
QY 1007 GCCTCGCTCACTGCTGATCGCGGTCAACGATCTGCTGTACGGGACCTCGCT 1059
DB 3931 GCCGCGCGCGGCTGCTGTGCGCGCGGATCGCGGAGTCTTGGACACCGGT 3879

RESULT 9
AA53491/C
ID AAX53491 standard; DNA; 114955 BP.
XX AC AAX53491;
XX AC AAX53491;
XX DT 05-JUL-1999 (first entry)
XX DE Human adenosine A1 receptor antisense oligonucleotide fragment.
XX KW Antisense oligonucleotide; multiple target; antisense treatment;
XX KW impaired respiration; inflammation; lung disease;
XX KW pulmonary vasoconstriction; inflammation; allergic rhinitis;
XX KW acute asthma; allergy; asthma; impeded respiration;
XX KW respiratory distress syndrome; pain; cystic fibrosis;
XX KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;
XX KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
XX KW colon cancer; breast cancer; lung cancer; pancreatic cancer;
XX KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
XX KW prostate cancer; ss.
XX OS Synthetic.
XX PN WO9913886-A1.
XX PD 25-MAR-1999.
XX PF 17-SEP-1998; 98WO-05019419.
XX PR 17-SEP-1997; 97US-0059160P.

```

09-JUN-1998; 98US-00093972.  
(UYEC-) UNIV EAST CAROLINA.  
Nyce JW;  
WPT; 1999-229400/19.  
New antisense oligonucleotides used in treatment of, e.g. pulmonary vasoconstriction.  
Disclosure; Page 37; 120pp; English.

The specification describes antisense oligonucleotides (AA52869-X55271) directed against at least 2 mRNAs selected from target genes, coding and non-coding regions of RNAs corresponding to target genes, gene initiation codons, genomic flanking regions, intron-exon borders, the 5'-end, the 3'-end and the juxta-section between coding and non-coding regions and all segments of RNAs encoding proteins associated with one or more diseases, conditions or mixtures. The antisense oligonucleotides may be derived from sequences AA55272-74. These multiple target oligonucleotides (specifically AA55180-271) can be used for the antisense treatment of diseases and conditions. Typical diseases and conditions are those associated with impaired respiration and inflammation, including lung diseases, pulmonary vasoconstriction, inflammation, allergic rhinitis, acute asthma, allergies, asthma, impaired respiration, respiratory distress syndrome, pain, cystic fibrosis, pulmonary hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukemias, lymphomas, carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic metastases, as well as all types of cancers which may metastasize or have metastasized to the lungs, including breast and prostate cancer

Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 0 U; 21427 Other;  
Query Match 4.7%; Score 55.2; DB 2; Length 114955;  
Best Local Similarity 31.6%; Pred. No. 0.32;  
Matches 298; Conservative 98; Mismatches 538; Indels 9; Gaps 4;

QY 234 CGTGAACCTCGGAGGCAACCGGATCGGCAACCACTCTCTGCCATCTCGGCATCTCAC 293  
Db 105259 CGCNNHNNNSCGGCCCGCGCGCGCGCGCCGCGCNCNNHNNNSCGGCCCGCGCGCGCGCC 105200  
QY 294 CCTTGACGAGCTGGACGGGACAGTCAACACAGCAGTGGGTCAACAGCTGGAGGCGCGGT 353  
Db 105199 NNHNNNSCGGCCCGCGCGCGCGCCVNNHNNNSCGGCCCGCGCGCGCGCCV 105140  
QY 354 CGGCCGCTGGCGGATCCACGCGGGTCTGTGGAAAGCGCTGGACGGAAGAGATCAG 413  
Db 105139 GNNHNNNSCGGCCCGCGCGCGCGCCVCGNNHNNNSCGGCCCGCGCGCGCGCGC 105080  
QY 414 CAACTGCCACGGCTCTGTCACACGATCTCGCGCGCGCGGATCGCGCGAAGCGATCAG 473  
Db 105079 CCVGGCNCNNHNNNSCGGCCCGCGCGCGCGCCVCGCNCNNHNNNSCGGCCCGCGCGC 105020  
QY 474 GCTCAACCGTCTCTTCGACGATGCTGCCCGCGCGAGCGCGAAGAGATGAAGTCTCT 533  
Db 105019 GCGCGCGCCVGGCCVNNHNNNSCGGCCCGCGCGCGCGCCVCGCNCNNHNNNSCG 104960  
QY 534 GAGCGAACCGGAGATCGGTC--GGCTTATACGGGGCTTCGCGGCACTGCGCACCGCTCG 592  
Db 104959 GCCCGCGCGCGCGCGCCVGGCVCNNHNNNSCGCCCGCGCGCGCGCGCGCGCGCG 104900  
QY 593 TCATGCTGCTGTGCGACCGGTCTGAGGTGGGTGAGGCGATCGGCTCTGCGCGCGCGCC 652  
Db 104899 CCVCGCNCNNHNNNSGGCCVCGCNCNNHNNNSVGGCCVCGCNCNNHNNNSCVGGCCVCGCNCNH 104840  
QY 653 GGGTCGACCTGCTCGCGCGCGCCCGGCTGACCGTCGTCGAGCAGCTCCAGGAGCTGG 712  
Db 104839 NNNSCVGGCCVCGCNCNNHNNNSCCCVGGCCVCGCNCNNHNNNSGCCCVCGCGCNCNH 104780  
QY 713 CCAGCACGGGAGAGCTGCTTCTTCAGTTCGCGAAGACCGCGAAGGCGCGCGCATCGGTCA 772

PS Disclosure; Page 37; 120pp; English.

XX The specification describes antisense oligonucleotides (AA552869-X55271) directed against at least 2 mRNAs selected from target genes, coding and non-coding regions of RNAs corresponding to target genes, gene initiation codons, genomic flanking regions, intron-exon borders, the 5'-end, the 3'-end and the juxta-section between coding and non-coding regions and all segments of RNAs encoding proteins associated with one or more diseases, conditions or mixtures. The antisense oligonucleotides may be derived from sequences AA55272-74. These multiple target oligonucleotides (specifically AA55180-271) can be used for the antisense treatment of diseases and conditions. Typical diseases and conditions are those associated with impaired respiration and inflammation, including lung diseases, pulmonary vasoconstriction, inflammation, allergic rhinitis, acute asthma, allergies, asthma, impeded respiration, respiratory distress syndrome, pain, cystic fibrosis, pulmonary hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukemias, lymphomas, e.g. colon cancer, breast cancer, lung cancer, pancreatic cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic metastases, as well as all types of cancers which may metastasize or have metastasized to the lungs, including breast and prostate cancer.

XX SQ Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 0 U; 21427 Other;

Query Match 4.7%; Score 55; DB 2; Length 114955;  
Best Local Similarity 30.7%; Pred. No. 0.34;  
Matches 249; Conservative 80; Mismatches 477; Indels 6; Gaps 2;  
QY 300 CGAGTGTGACGGGAGGTACCCAGAGTGGTCAACGACTGGAGCGCGGTGCGGCC 359  
DB 104645 BGGCGCGCCGNNNNNNCCGCGCGCGCGCGCGNNNNNNCCGCGCGCGCGCG 104704  
QY 360 GTGCGCGGA--GTCCAGCGGGGTCTCGGAAGCGCTGCGAGAGACGATCAGCAAC 417  
DB 104705 GCSNNNNCCGCGCGCGCGCGCGNNNNNNCCGCGCGCGCGCGCGNNNNCCG 104764  
QY 418 TGCCAGCGCTGTGCAACGATCTGCGCGCGCGCGATCGCGCGAAACGATCAGGTC 477  
DB 104765 BGGCGCGCGNNNNNNCCGCGCGCGCGCGNNNNNNCCGCGCGCGCGCGNNNN 104824  
QY 478 AACCGTGTCTTCGACGATCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 537  
DB 104825 CGCGCGCGCGNNNNNNCCGCGCGCGCGCGNNNNNNCCGCGCGCGCGCGCGCG 104884  
QY 538 GACCGGAGATCGTTCGCTTATCAGCGCGCTTCCGCGCGCTTCCGCGCGCGCGCG 597  
DB 104885 CCSNNNNCCG 104940  
QY 598 CTGCTGTGTGGACCGGTCTGAGTGGGTGAGCGATCGCGCGCGCGCGCGCGCG 657  
DB 104941 GCGCGCGCGCGCGCGCGCGCGCGNNNNNNCCGCGCGCGCGCGCGCGCGCGCG 105000  
QY 658 GACTGTGTGCGCGCGCGCGCGCGCGCTGACCGTGTGAGAGTCCAGAGTGGCGAG 717  
DB 105001 NDNNBGGCG 105060  
QY 718 ACGGAGAGTCTGCTTCCAGTCCCGAAGACCGAGCGCGCGCGCGCGCGCGCG 777  
DB 105061 GGGCGGNNNNNNCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105120  
QY 778 ACCACGAAGTCTGTCTTACCTCTTACCTCTTACCTCTTACCTCTTACCTCTTAC 837  
DB 105121 CG 105180  
QY 838 GTGTTCACCG 897  
DB 105181 CG 105240  
QY 898 AAGCGGTGCGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 957  
DB 105241 CCG 105300

QY 958 GCGCGCATCTGATTTTCTCGCGCGCTCGCTGTGCGGATCTCGCGCGCTCGGTAC 1017  
DB 105301 CGSNNNNCCGCGCGCGCGCGCGCGNNNNNNCCGCGCGCGCGCGCGNNNNNGC 105360  
QY 1018 TCGTGTGATCGGTTCAGGATCTGCTGTACGGGCGACCTGCTGAGAGGTTCAGGAGGG 1077  
DB 105361 GCGCGCGCGCGCGCGNNNNNNCCGCGCGCGCGCGCGCGNNNNNNCCGCGCGCGCGCG 105420  
QY 1078 ATCTCTCGCGCATCGAGGAGGCGATGGCGG 1109  
DB 105421 NDNNCCGCGCGCGCGCGCGNNNNNNCCGCGCGCGCGCGCGCGCGCG 105452

RESULT 11

ADA71938

ID ADA71938 standard; DNA; 2000 BP.

XX ADA71938;

XX ADA71938;

DT 20-NOV-2003 (first entry)

XX Rice gene, SEQ ID 5263.

XX Plant; bacterial infection; fungal infection; viral infection; rice;

XX gene; ds.

XX Oryza sativa.

XX WO2003000898-A1.

XX 03-JAN-2003.

XX 22-JUN-2001; 2001WO-IB001105.

XX 22-JUN-2001; 2001WO-IB001105.

XX (SYGN) SYNGENTA PARTICIPATIONS AG.

XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;

XX Katagiri F, Qian S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;

XX WPI; 2003-175290/17.

XX Identifying at least one gene involved in plant resistance or response to

XX pathogenic infection for conferring resistance or tolerance to a plant to

XX bacterial, fungal or viral infection by determining or detecting plant

XX gene expression.

XX Claim 27; SEQ ID NO 5263; 899pp; English.

XX The present invention relates to a method (M1) for identifying genes

XX involved in plant resistance or response to pathogenic infection. M1

XX comprises identifying a gene whose expression is significantly altered in

XX the incompatible interaction of plant gene expression relative to

XX expression of the gene in an uninfected plant, in a mutant plant that

XX does not express a gene associated with response to pathogenic infection,

XX or in a corresponding incompatible or compatible interaction. (M1) is

XX useful for conferring resistance to resistance or tolerance to a plant to

XX bacterial, fungal or viral infection. The present sequence was used to

XX illustrate the invention.

XX Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match 4.5%; Score 53.2; DB 7; Length 2000;

Best Local Similarity 10.5%; Pred. No. 0.77;

Matches 79; Conservative 353; Mismatches 311; Indels 11; Gaps 4;

QY 231 CGCGGTGAATCGAGGCGCAACCGGATCGGACACCTCTCGCCCATCTCGGCACT 290

DB 1 SRACGAWKSRCSWRGRRTMAGNMSCARGSSNRMRKMGMSKYRKCSGKRWTT 60

QY 291 CACCTTGTACGAGTGGAGCGGCGAGTCAACGAGCAGTGGGTCAACGACCTGGAGCGCG 350

Db 61 RRKSKWYSASSAS--GRTGSKWSSGSYSGKMKRYKSKRWGRGRGRGRGRGRGRGRGRGRGR 118  
Qy 351 CGTCGGCCCGTGGCGGAGTCCACGGGGGTCTGGGAAGCCGCTGGACGCGAGACGAT 410  
Db 119 YRRCARSGRAGGSGRMVGGKSRMSYMMWCYARGCGSKRKKSGKGGWGTCTRRGARGGS 178  
Qy 411 CAGCAACTGCACGGCTGTCTGCACACGATCTCGCGGGCGGATCGCGCGGAACGGAT 470  
Db 179 GWSGAKYKSGMSKRWMSRCRSGGRSAYRYGTSRYGTYKRYKMYTYSASRCMRA 238  
Qy 471 CAGGCTCAACCGTGTCTTCGACGATGTCGCCCGCGCGAGCCGAAAGAGATGAAGTT 530  
Db 239 YMTTSYWSAGSSYTCRSKRSMWMMKMRKMSRSGYVMSYKVMTCMTAYKKSYSYRW 298  
Qy 531 CCGTAGCG-ACCGGAGATCGGTCTGCTATACGCGGCTTCGCGCGACTGCGGACCG- 588  
Db 299 CYMGGGWRGATRYWGRGMSRMAMMYKMYRYGKMGKRWGAWGRMMSMCRWSKA 358  
Qy 589 -CTCGTCATGCTGCTGGTGGCAGCCGCTGTAGGTGGGTGAGCGGATCGGCTGGCGC 647  
Db 359 CYMWRWRMWRTHRRRWAWSKSRSTSRKKRCKMRKRYKRYGYSRMSCKYRWMKRC 418  
Qy 648 CGCGCGGTGACCTGCTCGCGCGCGCGCGCGGTGACCGTGTGAGCAGCTCAAGGA 707  
Db 419 RSGRAWKMGRCGCTCRMSYGMWRKWSKRWASKYKMSRMRYRWRKKCSRTTMMGTR 478  
Qy 708 GCTGGCCAGCAGCGGAGAGCTGCTCTCCA-----GTGCGCGAAGACCGCGAAGCGCG 761  
Db 479 GGMGTGRCRYKRSKMRKRRRWRGRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYR 538  
Qy 762 GCGACGGTCACTTCCACCAAGTTCGCTTACTGCTTACCGCACTCATCGCGGAAA 821  
Db 539 KARCWYRGYVWAGWMMKRYKRYMYKMYKMYKMYKMYKMYKMYKMYKMYKMYKMYKMY 598  
Qy 822 GAAAGTACGAGGTGCTGTCACCGCGCGGAAAGCGGGATGTTAAGACCGCGCAATT 881  
Db 599 AXMKRSKMSAWSKMSRKRCKKASRKSARVAMMGMTSGSRMSRWKSYTCYWRKW 658  
Qy 882 CGCGCGGATCTGGTCAAGCGTGGAGGAGCGCGGCTTCGCGGCTTACGCAATCAAGA 941  
Db 659 GSKSTCTWYMSKYVYAKYSGYWRYYRACMYMRWYRYRYRYRYRYRYRYRYRYRYRY 718  
Qy 942 TCTGGGACACATCAGCGCGGATCTGATTCT 975  
Db 719 AMTGMYSGRYWTSWYKCKSKRYSMWYMSW 752

RESULT 12  
AAL61190  
ID AAL61190 standard; DNA; 504 BP.  
XX AAL61190;  
AC AAL61190;  
XX 22-SEP-2003 (first entry)  
DT Actinosynnema pretiosum O-methyltransferase gene.  
DE Actinosynnema pretiosum O-methyltransferase gene.  
XX Maytansinoid; ansamitocin; antitumour; O-methyltransferase; enzyme; gene;  
KW ds.  
XX Actinosynnema pretiosum.  
OS WO2003045312-A2.  
PN 05-JUN-2003.  
XX 21-NOV-2002; 2002WO-US037547.  
XX 21-NOV-2001; 2001US-0332158P.  
XX (UNIW ) UNIV WASHINGTON.  
PA Floss HG, Yu T, Leistner E;  
PI

XX WPI; 2003-493374/46.  
XX Novel maytansinoid produced by bacterial host cell transformed with  
PT expression vector comprising open reading frame from ansamitocin gene  
PT cluster I of Actinosynnema pretiosum, useful as antitumor agent.  
XX Disclosure; Page 80-81; 160pp; English.  
XX The invention relates to maytansinoid produced by bacterial host cell  
CC transformed with expression vector comprising open reading frame from  
CC ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is  
CC useful as a potent antitumour agent. The present sequence is A. pretiosum  
CC ansamitocin gene cluster I O-methyltransferase gene  
XX Sequence 504 BP; 57 A; 192 C; 190 G; 65 T; 0 U; 0 Other;  
SQ

Query Match 4.5%; Score 53; DB 7; Length 504;  
Best Local Similarity 46.6%; Pred. No. 0.86;  
Matches 170; Conservative 0; Mismatches 195; Indels 0; Gaps 0;  
Qy 404 AGACGATCAGCAACTGCCACGGCCTGCTGCACACGATCTGCGGCGCGGATCGCGCGA 463  
Db 116 AGCTGAGCCCAAGTGGCCGCGATCGCCGCGCGCACTGGAGCGCGGGCGTCCGG 175  
Qy 464 AACGATCAGGCTCAACCCGTGCTTTCGACGATGCTGCCCGCGCGAGCGCGAAGAGA 523  
Db 176 ACCGATCGAGGTGCTGTCGGCGACGCCGCGGCGAGCCTCGCGGGCTGACCGCGGAGC 235  
Qy 524 TGAAGTTCCTGAGCGACCCGCGAGATCGTTCGGCTTATACGGCGCTTCCGCGCACTGGC 583  
Db 236 TCGCGGCGCGGTGCACTGGTGTTCGTGACGCGGACAAGCGCGGTACTCCCGCTACT 295  
Qy 584 GACCGCTGCTCATCTGCTGCTGCTGCGACCGGTCTGAGTGGGTGAGCGGATCGGCTGC 643  
Db 296 ACAGACTGGCGTTCGAGTGTGCGCCCGCGGGCTGGTCTGTCGTCGACACACACCTGT 355  
Qy 644 GCGCGCGCGGTGCACTGCTGCGCGCGCGCGCGCTGACCGTGTGAGCGAGCTCC 703  
Db 356 TCTCGGCGGGTTCGCGACCCCGCGGTGACCCCGGACACCGAGGGGTTCGGGAGC 415  
Qy 704 AGGAGTGGCGACGACGCGGAGAGTCTGTTTCCAGTTCGCGGAAAGACCGCGAGGCGCGC 763  
Db 416 TCAACCGCAGATCGCGCGCGACGACCGGTTCAGGCGGTGCTGCTCGCGCACGCGGAGC 475  
Qy 764 GCAG 768  
Db 476 GCATG 480

RESULT 13  
AAL61224/c  
ID AAL61224 standard; DNA; 82746 BP.  
XX AAL61224;  
AC AAL61224;  
XX 22-SEP-2003 (first entry)  
DT Actinosynnema pretiosum ansamitocin biosynthetic gene cluster I.  
DE Actinosynnema pretiosum.  
XX Maytansinoid; ansamitocin; antitumour; gene; ds.  
XX Actinosynnema pretiosum.  
OS WO2003045312-A2.  
PN 05-JUN-2003.  
XX 21-NOV-2002; 2002WO-US037547.  
XX 21-NOV-2001; 2001US-0332158P.  
XX (UNIW ) UNIV WASHINGTON.  
PA



```
FT FT CDS complement(20904..22094)
FT FT /*tag= t
FT FT /product= "quinone-forming hydroxylase"
FT FT /note= "gene T (specifically claimed)"
FT FT complement(22505..22179)
FT FT /*tag= u
FT FT /product= "not specified"
FT FT /note= "gene U"
XX XX EP806480-A2.
XX XX 12-NOV-1997.
XX XX 02-MAY-1997; 97EP-00107329.
XX XX 07-MAY-1996; 96US-0016753P.
XX XX (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX XX Reeves CD, Soliday CL;
XX XX WPI; 1997-538619/50.
XX XX P-PSDB: AAW34159, AAW34200, AAW34201, AAW34202, AAW34203, AAW34204,
XX XX AAW34205, AAW34206, AAW34207, AAW34208, AAW34209, AAW34210, AAW34211,
XX XX AAW34212, AAW34213, AAW34214, AAW34215, AAW34216, AAW34217, AAW34218,
XX XX AAW34219.
XX XX Streptomyces frenolicin gene cluster - useful for producing recombinant
XX XX frenolicin antibiotics.
XX XX Claim 1; Page 40-60; 66pp; English.
XX XX This DNA sequence comprises the Streptomyces frenolicin gene cluster
XX XX containing specifically claimed coding sequences (genes A-U) that
XX XX respectively encode 21 proteins (see AAW34199-219) involved in frenolicin
XX XX synthesis. The genes can be divided into 5 subclusters: (1) genes A, B,
XX XX C, D and F encode an efflux pump; (2) genes H, I, J and K encode butyrate
XX XX starter synthases; (3) genes L, M and N encode polyketide synthases (PKS)
XX XX ; (4) genes O, P, Q and R encode a hemiketalase, a ketoreductase and
XX XX cyclases/dehydroases; and (5) genes S and T encode a keto/enoyl reductase
XX XX and a hydrolase. Also claimed are vectors, host cells (especially a
XX XX Streptomyces sp., particularly Streptomyces roseofulvus), and the encoded
XX XX proteins. Cells transformed using the above sequence can be cultured to
XX XX produce frenolicins or frenolicin precursors. The precursors can be
XX XX converted to frenolicins by chemical or other methods. The frenolicins
XX XX can be oxidised to frenolicin B, an antibiotic used as an anticoccidial
XX XX agent. The frenolicins can be used as animal feed additives
XX XX
XX XX Sequence 24379 BP; 3077 A; 9792 C; 8499 G; 3011 T; 0 U; 0 Other;
SQ
Query Match 4.5%; Score 52.8; DB 2; Length 24379;
Best Local Similarity 52.2%; Pred. No. 0.89;
Matches 117; Conservative 0; Mismatches 107; Indels 0; Gaps 0;
QY 945 GCGGCACACTACACGGCGGATCCTGATTCTCGCGGGCTCGCTGTCGGCGATCTCCCG 1004
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3577 GCGGTACGGGACGTGGCGCTCGCGGGCGGGGGCCAGGCTGCTGCTCGCG 3636
QY 1005 CCGCTCGGTACTGTCGTATCGCGGTACGATCTGTGTACGGCACCTGCGGTGAGGA 1064
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3637 CCGCTGCGCAACGGCTCGGTGGCGCCCTGGCCCTGCTCGCGGCTGCTGCGGCACT 3696
QY 1065 GGTGCACAGGGGATCTTCGCGGGCATCGAGGAGCGATGCGCGCTCGGGTGAGGA 1124
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3697 CTTCCCGGGGCGCTCTGTGTGGAGACGTTCTTCGACAGGGGCGCATCGGCGCTCGC 3756
QY 1125 CTTGAGGCGGAACTCGACGAGGACTGACGAGCTGTGTGCGG 1168
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3757 CATGACGCGCTGCTACCTACGAGGACTGCGGCTGTCGACGGCG 3800
RESULT 15
AAV25925
```

AAV25925 standard; cDNA; 24379 BP.

AAV25925;

15-JUL-1998 (first entry)

Streptomyces roseofulvus frenolicin gene cluster.

Streptomyces roseofulvus; frenolicin gene cluster; frenolicin B; antibiotic; ss.

Streptomyces roseofulvus.

Location/Qualifiers  
636..2948  
/\*tag= a  
/transl\_except= (pos:636..638,aa:Met)  
/note= "encodes protein given in AAW55800"  
2945..3916  
/\*tag= b  
/transl\_except= (pos:2945..2947,aa:Met)  
/note= "encodes protein given in AAW55801"  
4020..4844  
/\*tag= c  
/transl\_except= (pos:4020..4022,aa:Met)  
/note= "encodes protein given in AAW55802"  
4841..6415  
/\*tag= d  
/note= "encodes protein given in AAW55803"  
6533..7183  
/\*tag= e  
/transl\_except= (pos:6533..6535,aa:Met)  
/note= "encodes protein given in AAW55804"  
7344..8897  
/\*tag= f  
/note= "encodes protein given in AAW55805"  
9184..10012  
/\*tag= g  
/transl\_except= (pos:9184..9166,aa:Met)  
/note= "encodes protein given in AAW55806"  
complement(10105..10621)  
/\*tag= h  
/transl\_except= (pos:10621..10619,aa:Met)  
/note= "encodes protein given in AAW55807"  
complement(10618..11628)  
/\*tag= i  
/transl\_except= (pos:11628..11626,aa:Met)  
/note= "encodes protein given in AAW55808"  
11809..12066  
/\*tag= j  
/note= "encodes protein given in AAW55809"  
complement(12154..13209)  
/\*tag= k  
/transl\_except= (pos:13209..13207,aa:Met)  
/note= "encodes protein given in AAW55810"  
13409..14686  
/\*tag= l  
/transl\_except= (pos:13409..13411,aa:Met)  
/note= "encodes protein given in AAW55811"  
14767..16047  
/\*tag= m  
/note= "encodes protein given in AAW55812"  
16120..16371  
/\*tag= n  
/note= "encodes protein given in AAW55813"  
complement(16453..16935)  
/\*tag= o  
/transl\_except= (pos:16935..16933,aa:Met)  
/note= "encodes protein given in AAW55814"  
17088..18903  
/\*tag= p  
/note= "encodes protein given in AAW55815"  
17903..18898

Job time : 705.638 secs

```
FT      /tag= q
FT      /note= "encodes protein given in AAW55816"
FT      1895. .19839
FT      /tag= r
FT      /note= "encodes protein given in AAW55817"
FT      complement(1990. .20907)
FT      /tag= g
FT      /transl_except= (pos:20907. .20905,aa:Met)
FT      /note= "encodes protein given in AAW55818"
FT      complement(20904. .22094)
FT      /tag= t
FT      /transl_except= (pos:22040. .22038,aa:Arg)
FT      /transl_except= (pos:22037. .22035,aa:Arg)
FT      /transl_except= (pos:22034. .22032,aa:Thr)
FT      /transl_except= (pos:22031. .22029,aa:Iys)
FT      /transl_except= (pos:22029,aa:Iys)
FT      /note= "encodes protein given in AAW55819"
XX
XX      JP10094395-A.
XX
XX      14-APR-1998.
XX
XX      07-MAY-1997; 97JP-00116652.
XX
XX      07-MAY-1996; 96US-0016753P.
XX      04-APR-1997; 97US-0042935P.
XX      (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
XX      WPI; 1998-279231/25.
XX      P-PSDB; AAW55800, AAW55801, AAW55802, AAW55803, AAW55804, AAW55805,
XX      AAW55806, AAW55807, AAW55808, AAW55809, AAW55810, AAW55811, AAW55812,
XX      AAW55813, AAW55814, AAW55815, AAW55816, AAW55817, AAW55818, AAW55819.
XX
XX      Frenolicin gene cluster - useful as an antibiotic.
XX
XX      Claim 2; Page 36-44; 50pp: Japanese.
XX
XX      The present sequence represents a frenolicin gene cluster from
XX      Streptomyces roseofulvus. The present invention describes: (1) a vector
XX      containing the frenolicin gene cluster DNA sequence operably connected to
XX      an expression control sequence; (2) a host cell transformed by the above
XX      vector; (3) a protein coded by the above DNA sequence; (4) a method for
XX      the preparation of frenolicin or a biosynthetic intermediate for it in
XX      which the above cell is cultured and frenolicin or its biosynthetic
XX      intermediate is isolated from the culture or the cell; (5) a method for
XX      the preparation of frenolicin B by oxidising frenolicin, and (6) a method
XX      for the preparation of a feed composition by mixing frenolicin with other
XX      components. Frenolicin B is useful as an antibiotic
XX
XX      Sequence 24379 BP; 3077 A; 9792 C; 8499 G; 3011 T; 0 U; 0 Other;
SQ
Query Match 4.5%; Score 52.8; DB 2; Length 24379;
Best Local Similarity 52.2%; Pred. No. 0.89;
Matches 117; Conservative 0; Mismatches 107; Indels 0; Gaps 0;
QY 945 GCGGCACATCAGCGCGGATCTGATTTCTGCGCGGCTCGCTGCTCGCGATCTCCG 1004
DB 3577 GCGGTACCGGAGCTGCGCGCTCGCGCGCGCGCGCGCGCGCGCTGCTGCTCGCGG 3636
QY 1005 CCGCCTCGGTCACTGCTGATCGCGGTACGGAATCTGTTACGGGACCTGCGTGAGGA 1064
DB 3637 CGCGCTCGGCAACGCTCGGTGCGCGCTGCGCTGCTGCGGCGCTGCTGCGGCAACT 3696
QY 1065 GTTCGACAGGGATCTTCGCGCGATCGAGGAGGATGCGCGGCTCGGGCTGAGGA 1124
DB 3697 CCTCGCGGGGCGGTCTCTGTTGAGACCGTCTTCGCGAGGGCGGATCGGGCGCTCGC 3756
QY 1125 CTTGAGCGGGAATCTCGACAGGAGTTCGACGACGCTGTTGGCCG 1168
DB 3757 CATGACGCGGTACCTACCGACCTGCCGCTCGTGCAGGGGG 3800
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2004, 08:50:04 ; Search time 5413.76 Seconds  
(without alignments)  
9439.175 Million cell updates/sec

Title: US-09-855-340A-1

Perfect score: 1179

Sequence: 1 ggtgtagcagaagaacgg.....tgtggcgcagcagcatga 1179

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*

30: em.htg.hum.\*

31: em.htg.inv.\*

32: em.htg.other.\*

33: em.htg.mus.\*

34: em.htg.pln.\*

35: em.htg.rod.\*

36: em.htg.nam.\*

37: em.htg.vrt.\*

38: em.sy.\*

39: em.htgo.hum.\*

40: em.htgo.mus.\*

41: em.htgo.other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1179	100.0	1179	6	AX338970	Sequence
2	1179	100.0	2025	7	AY150027	Bacterioph
3	1176	99.7	1176	6	AX196104	Sequence
4	66.4	5.6	347660	1	AF002994	AP002994
5	65.8	5.6	300800	1	AP005036	Streptomy
6	65.6	5.6	125020	9	AF429315	Streptomy
7	62.4	5.3	125020	9	AF429315	Homo sapi
8	61.6	5.2	2404	6	AB1401	Sequence
9	61.6	5.2	2404	7	AF030986	Streptomy
10	60.4	5.1	2000	6	AX655393	Streptomy
11	59.8	5.1	231001	2	AC130981	Rattus no
12	59.8	5.1	258319	2	AC130981	Rattus no
13	59.4	5.0	1094	11	PM7G11B	AL658196
14	59.4	5.0	3689	12	AF260831	Integrati
15	59.4	5.0	56425	7	AF165214	Bacterioph
16	57.8	4.9	2330	24	AVBHINT	Actinoph
17	57.8	4.9	49220	7	AF320035	Bacterioph
18	56.6	4.8	1161	6	AR339658	Sequence
19	56.6	4.8	1161	6	AX275016	Sequence
20	56.6	4.8	11188	6	AR339644	Sequence
21	56.6	4.8	11188	6	AX275002	Sequence
22	56.6	4.8	132544	1	AF521085	Streptomy
23	56.6	4.8	231000	1	SCO939105	Streptomy
24	56.4	4.8	67480	7	AY129334	Mycobacte
25	56.4	4.8	293050	1	SCO939116	Mycobacte
26	56.2	4.8	156218	2	AP004994	Oryza sat
27	55.8	4.7	27541	6	AX211706	Sequence
28	55.8	4.7	123580	1	AF263912	Streptomy
29	55.8	4.7	125401	6	AX211739	Sequence
30	55.4	4.7	65351	2	AC139773	Homo sapi
31	55.4	4.7	245476	2	AC098496	Rattus no
32	54.8	4.6	230278	14	MCU68299	Mouse cyto
33	54.2	4.6	303550	1	SCO939118	Streptomy
34	54.2	4.6	321250	1	SCO939111	Streptomy
35	53.8	4.6	1603	5	AF071875	Gallus ga
36	53.8	4.6	311000	1	SCO939122	Streptomy
37	53.6	4.5	300425	1	AP005041	Streptomy
38	53.2	4.5	2000	6	AX655393	Sequence
39	53.2	4.5	39250	1	SVI011500	Streptomy
40	53.2	4.5	132544	1	AF521085	Streptomy
41	53	4.5	575	1	MSGINT	Mycobacteri
42	53	4.5	82746	1	AF453501	Actinosyn
43	52.8	4.5	25306	1	AF058302	Streptomy
44	52.6	4.5	308050	1	SCO939124	Streptomy
45	52.2	4.4	300425	1	AP005044	Streptomy

## ALIGNMENTS

RESULT 1  
AX338970  
LOCUS AX338970  
DEFINITION Sequence 1 from Patent WO0187936.  
ACCESSION AX338970  
VERSION AX338970.1  
KEYWORDS GI:18129106  
SOURCE Micromonospora carbonacea  
ORGANISM Micromonospora carbonacea  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Micromonosporineae; Micromonosporaceae; Micromonospora.  
REFERENCE 1  
Hosted, T.J. and Horan, A.C.  
AUTHORS Isolation of Micromonospora carbonacea var africana pm1 pl integrate  
TITLE and use of integrating function for site-specific integration into

Pred. No. is the number of results predicted by chance to have a

Micromonospora halophitica and Micromonospora carbonacea chromosome  
Patent: WO 0187936-A 1 22-NOV-2001;  
SCHERING CORPORATION (US)

FEATURES  
source  
1..1179  
/organism="Micromonospora carbonacea"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:47853"

ORIGIN

Query Match 100.0%; Score 1179; DB 6; Length 1179;  
Best Local Similarity 100.0%; Pred. No. 4.3e-159;  
Matches 1179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGTGATCGAGAGAACGGGCGGCTCTACCCATTCGGGACCTCTGTCGGGGTAAAAAG 60  
Db 1 GTGTGATCGAGAGAACGGGCGGCTCTACCCATTCGGGACCTCTGTCGGGGTAAAAAG 60  
Qy 61 GTACCAATTCAGACCGGTTATCGAGAGACGAGCCCAAGAAATCGGATGTCAGTTTC 120  
Db 61 GTACCAATTCAGACCGGTTATCGAGAGACGAGCCCAAGAAATCGGATGTCAGTTTC 120  
Qy 121 CGTCCGAGCAGTTGACGGGCAACGGGCTCATCGCGCGCGGGGCTCAGATTACCTCCGC 180  
Db 121 CGTCCGAGCAGTTGACGGGCAACGGGCTCATCGCGCGCGGGGCTCAGATTACCTCCGC 180  
Qy 181 GATTTCGTGGGAGTGGTGGCGAGCTACGAAAGACGCTGAAACCGACCGCGTGAAC 240  
Db 181 GATTTCGTGGGAGTGGTGGCGAGCTACGAAAGACGCTGAAACCGACCGCGTGAAC 240  
Qy 241 TCGGAGGCAACCGGATTCGCAACACCTCTCTGCCCATCTCGGCACTCTACCTCTTGC 300  
Db 241 TCGGAGGCAACCGGATTCGCAACACCTCTCTGCCCATCTCGGCACTCTACCTCTTGC 300  
Qy 301 GAGCTGAGCGGAGTGCACCAAGAGTGGTCAACGACTGAGGCGCGGCTCGGCCG 360  
Db 301 GAGCTGAGCGGAGTGCACCAAGAGTGGTCAACGACTGAGGCGCGGCTCGGCCG 360  
Qy 361 TGGCGGAGTCCAGCGGGGTCTCGAAGCGCTGCGGAGGAGGAGTCAAGTCACTG 420  
Db 361 TGGCGGAGTCCAGCGGGGTCTCGAAGCGCTGCGGAGGAGGAGTCAAGTCACTG 420  
Qy 421 CAGCGCTGTGACACGCTCTCGGCGGGGCTGCGGCGGAGTCCGCGGAAACGAGTCAAC 480  
Db 421 CAGCGCTGTGACACGCTCTCGGCGGGGCTGCGGCGGAGTCCGCGGAAACGAGTCAAC 480  
Qy 481 CCGTGTCTTCGACGATCTGCCCCCGGCGGAGCGGAGGAGTCAAGTCTGAGCGAC 540  
Db 481 CCGTGTCTTCGACGATCTGCCCCCGGCGGAGCGGAGGAGTCAAGTCTGAGCGAC 540  
Qy 541 CCGGAGATCGTTCGGCTTATCAGGGGCTTCCCGCGCACTGCGGACCGCTCTCATGCTG 600  
Db 541 CCGGAGATCGTTCGGCTTATCAGGGGCTTCCCGCGCACTGCGGACCGCTCTCATGCTG 600  
Qy 601 CTGCTGCGACCGGCTCTGAGTGGGTGAGCGGATCGGCTGCGCGCGCGGCTCGAC 660  
Db 601 CTGCTGCGACCGGCTCTGAGTGGGTGAGCGGATCGGCTGCGCGCGCGGCTCGAC 660  
Qy 661 CTGCTGCGCGCGGCGCGGCTGACCGTCTGAGCAGCTCCAGGAGTGGCCAGCAG 720  
Db 661 CTGCTGCGCGCGGCGCGGCTGACCGTCTGAGCAGCTCCAGGAGTGGCCAGCAG 720  
Qy 721 GGAGAGTCTCTTCCAGTCCGCEAAGACCGGAGGCGCGGCGACCGTCTGAGTTTCAAC 780  
Db 721 GGAGAGTCTCTTCCAGTCCGCEAAGACCGGAGGCGCGGCGACCGTCTGAGTTTCAAC 780  
Qy 781 ACAGAGTCTCTTCTACTCTTACGCTTACGCTCATCTCCGGAAGAAAAGTGCAGGTCGTG 840  
Db 781 ACAGAGTCTCTTCTACTCTTACGCTTACGCTCATCTCCGGAAGAAAAGTGCAGGTCGTG 840  
Qy 841 TTACCGCGCGGAGGCGGAGTGGTAAGACGCGCAATTTCCGCGGATCTGGGTCAAG 900  
Db 841 TTACCGCGCGGAGGCGGAGTGGTAAGACGCGCAATTTCCGCGGATCTGGGTCAAG 900

Qy 901 GCCTGCGAGAACCGGGCTTCGGGCTTACGATTCAGATTCGGGACACTCAGCG 960  
Db 901 GCCTGCGAGAACCGGGCTTCGGGCTTACGATTCAGATTCGGGACACTCAGCG 960  
Qy 961 GCGATCTGATTTCTCGCGGGGCTCGGCTGTGGCGATCTCCGGCGCTTCGGTCACTCG 1020  
Db 961 GCGATCTGATTTCTCGCGGGGCTCGGCTGTGGCGATCTCCGGCGCTTCGGTCACTCG 1020  
Qy 1021 TCGATCGCGTACGAGATCTGCTGACGGGACCTCGGAGGAGTCCAGCGGGGATC 1080  
Db 1021 TCGATCGCGTACGAGATCTGCTGACGGGACCTCGGAGGAGTCCAGCGGGGATC 1080  
Qy 1081 CTGCGGCGATCGAGAGGCGATGCGCGCTCGGCTCGGCTGAGGACCTGGAGCGGAATC 1140  
Db 1081 CTGCGGCGATCGAGAGGCGATGCGCGCTCGGCTCGGCTGAGGACCTGGAGCGGAATC 1140  
Qy 1141 GACGAGGAGTACGAGCGGCTGTTGGCGGACGAGCATGA 1179  
Db 1141 GACGAGGAGTACGAGCGGCTGTTGGCGGACGAGCATGA 1179  
RESULT-2  
AY150027 2025 bp DNA linear PHG 02-SBP-2003  
DEFINITION Bacteriophage pMLP1 att/int region.  
ACCESSION AY150027  
VERSION AY150027.1 GI:28630421  
KEYWORDS Bacteriophage pMLP1  
SOURCE Bacteriophage pMLP1  
ORGANISM Viruses.  
REFERENCE 1 (bases 1 to 2025)  
AUTHORS Alexander,D.C., Devlin,D.J., Hewitt,D.D., Horan,A.C. and Hosted,T.J.  
TITLE Development of the Micromonospora carbonacea var. africana ATCC 39149 bacteriophage pMLP1 integrase for site-specific integration in Micromonospora spp  
JOURNAL Microbiology (Reading, Engl.) 149 (Pt 9), 2443-2453 (2003)  
PUBMED 12949170  
REFERENCE 2 (bases 1 to 2025)  
AUTHORS Hosted,T.J. Jr., Alexander,D.C. and Hewitt,D.D.  
TITLE Direct Submission  
JOURNAL Submitted (12-SBP-2002) New Lead Discovery, Schering-Plough Research, 2015 Galloping Hill Road, Kenilworth, NJ 07033, USA  
FEATURES  
source  
1..2025  
/organism="Bacteriophage pMLP1"  
/mol\_type="genomic DNA"  
/specific\_host="Micromonospora sp. ATCC 39149"  
/db\_xref="taxon:219292"  
/note="host is deposited in ATCC as Micromonospora carbonacea var. africana Waitz et al."  
misc\_feature  
1..2025  
/note="att/int region"  
197..622  
/note="xis"  
/codon\_start=1  
/transl\_table=11  
/product="excisionase"  
/protein\_id="AA046044.1"  
/db\_xref="GI:28630422"  
/translation="MRNTPGLGRGTAAVYLTARERAGLTKSELARRIKQDRATGRW  
EDGNRPDADLVARVAQLGLDLDEALAAAGLRPGVTPPATPTMDLDEIELVTRDP  
KLDMDKRIALLILERRERDKAAIETKRLIDLFRS"  
628..1806  
/gene="intX"  
628..1806  
/gene="intM"  
/codon\_start=1  
/transl\_table=11  
/product="integrase"  
/protein\_id="AA046045.1"

```
/db_xref="GI:28630423"
/translation="MMIKNGPVYRIRDLVRGKKVTIQTGYPTKTSKXNMMVQFRAEQ
LQGNALMPRGQITLADFYGEWMPSEYKTLKPTAVNSEGNIRNHLILTLDEL
DGOVTOQWVNDLEAGVPWPESTRGRKPLAAKTSINCHGLLHTICGAAIAKRLIN
PCSTMLPREPEKMFSLDPEIGRLITALPHWRPLVLLVATGLRGEALGRAGR
VDLLAARPLTVVEQLQELASTGELVFQSPKAKRRVTSFTTKVALLITPLIAGKKS
DEVVTFAPKGMVTRNFRIRIWKACEAGLPLGRIRHDLRHTHAALISAGRPISAI
RLGHSSIAVTDLLYGLHREVEDEGILAAIEAMAGVRAEDLEALDELDLVLADAA
"
stem_loop      join(1884..1895,1900..1911)
misc_feature    1925..1986
                 /note="attp element"
stem_loop      join(1969..1986,1991..2010)
ORIGIN
Query Match      100.0%; Score 1179; DB 7; Length 2025;
Best Local Similarity 100.0%; Pred. No. 3.8e-159;
Matches 1179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGTGGATCGAAGAACGGGCCGCTACCGCATTCGGGACCTGTTCCGGTAAAAG 60
DB 628 GTGTGGATCGAAGAACGGGCCGCTACCGCATTCGGGACCTGTTCCGGTAAAAG 687
QY 61 GTCACATTACAGACCGGTTATCCGACGAAGACAGCGCCAAAGATCCGATGTCAGTTC 120
DB 688 GTCACATTACAGACCGGTTATCCGACGAAGACAGCGCCAAAGATCCGATGTCAGTTC 747
QY 121 CGTGGGACAGTTGAGGGCAACGGCTCATGCGCGCGGGCGGTCAATACCTTCGCC 180
DB 748 CGTGGGACAGTTGAGGGCAACGGCTCATGCGCGCGGGCGGTCAATACCTTCGCC 807
QY 181 GATTTCGTGGGAGTGGTGGCGGAGCTACGAAAGACAGCTCAAAACCGCCGCTGGAAC 240
DB 808 GATTTCGTGGGAGTGGTGGCGGAGCTACGAAAGACAGCTCAAAACCGCCGCTGGAAC 867
QY 241 TCGAGGGCAACCGGATCCGAACCACTCTCGGCCATCTCGGCCATCTCAACCTTCAC 300
DB 868 TCGAGGGCAACCGGATCCGAACCACTCTCGGCCATCTCGGCCATCTCAACCTTCAC 927
QY 301 GAGCTGGACGGCAGTCAACGACAGTGGTCAACGACCTGGAGCGGGCGGTTCGCCCG 360
DB 928 GAGCTGGACGGCAGTCAACGACAGTGGTCAACGACCTGGAGCGGGCGGTTCGCCCG 987
QY 361 TCGCGGAGTCCACGCGGGTCTCGGAAGCGCTGGCAGCGAAGACGATCAGCAACTGC 420
DB 988 TCGCGGAGTCCACGCGGGTCTCGGAAGCGCTGGCAGCGAAGACGATCAGCAACTGC 1047
QY 421 CACGCGCTCTGCACACGATCTCGCGCGGGCGATCGCGCGGAACCGATCAGGCTCAAC 480
DB 1048 CACGCGCTCTGCACACGATCTCGCGCGGGCGATCGCGCGGAACCGATCAGGCTCAAC 1107
QY 481 CCGTGCTCTTCGACGATCTGCCCGCGCGAGCCGAAAGAGATGAAGTTCCTGAGCGAC 540
DB 1108 CCGTGCTCTTCGACGATCTGCCCGCGCGAGCCGAAAGAGATGAAGTTCCTGAGCGAC 1167
QY 541 CCGGAGATCGTGGGTTATACCGCGGTTTCGCCGCACTGGCGACCGCTCGTATGCTG 600
DB 1168 CCGGAGATCGTGGGTTATACCGCGGTTTCGCCGCACTGGCGACCGCTCGTATGCTG 1227
QY 601 CTGTTGGCGACCGGTCGTAGGTGGGTGAGGGGATCGGCCTCGCGCGCGCGGGTGCAC 660
DB 1228 CTGTTGGCGACCGGTCGTAGGTGGGTGAGGGGATCGGCCTCGCGCGCGGGTGCAC 1287
QY 661 CTGCTCGCGCGCGCGCGGCTGACCGTCTGTCGAGCAGCTCCAGGAGCTGGCCAGCAG 720
DB 1288 CTGCTCGCGCGCGCGCGGCTGACCGTCTGTCGAGCAGCTCCAGGAGCTGGCCAGCAG 1347
QY 721 CGAGAGCTCGTTCAGTTCGCGGAAGACCGGAGGGCGCGCCACCGTCAGTTTCACC 780
DB 1348 GAGAGCTCGTTCAGTTCGCGGAAGACCGGAGGGCGCGCCACCGTCAGTTTCACC 1407
QY 781 ACGAAGTCTGCTACTGCTTACGCCACTCATCGCCGGAAGAAAGATGACGAGTCTGT 840
```

```
Db 1408 ACGAAGTCGCTACTGCTTACGCCACTCATCCCGGAAAGAAAGTACGAGGTCGTG 1467
QY 841 TTTCACCGCGCCGAAAGCGGGATGTAAGACGCGCAATTTCCGCGGATCTGGTCAAG 900
DB 1468 TTTCACCGCGCCGAAAGCGGGATGTAAGACGCGCAATTTCCGCGGATCTGGTCAAG 1527
QY 901 CGGTGCGAGGAAGCGGGCTTCCGGGTTACGCAATTCACGATCTGCGGCACATCAGCG 960
DB 1528 CGGTGCGAGGAAGCGGGCTTCCGGGTTACGCAATTCACGATCTGCGGCACATCAGCG 1587
QY 961 CGCATCTCTGATTTCTCGCGCGGCTCCGCTGTCGCGGATCTCCCGCGGCTCGGTCACTCG 1020
DB 1588 CGCATCTCTGATTTCTCGCGCGGCTCCGCTGTCGCGGATCTCCCGCGGCTCGGTCACTCG 1647
QY 1021 TCGATCGCGGTTCACGATCTGCTGTACGGGCACCTGCGGTGAGGAGTTCGACGAGGATC 1080
DB 1648 TCGATCGCGGTTCACGATCTGCTGTACGGGCACCTGCGGTGAGGAGTTCGACGAGGATC 1707
QY 1081 CTCGCGGCGATCGAGGAGCGATGCGCGGCTCGCGGCTGAGGACCTGAGGCGGAACTC 1140
DB 1708 CTCGCGGCGATCGAGGAGCGATGCGCGGCTCGCGGCTGAGGACCTGAGGCGGAACTC 1767
QY 1141 GACGAGGAGCTGACGAGACGCTGTTGCCCGGACGACGATGA 1179
DB 1768 GACGAGGAGCTGACGAGACGCTGTTGCCCGGACGACGATGA 1806

RESULT 3
AX196104 LOCUS 1176 bp DNA linear PAT 28-AUG-2001
DEFINITION Sequence 176 from Patent WO0151639.
ACCESSION AX196104
VERSION AX196104.1 GI:15386336
KEYWORDS
SOURCE Micromonospora carbonacea
ORGANISM Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
REFERENCE 1
AUTHORS Hosted, T.J., Horan, A.C. and Wang, T.X.
TITLE Evernimycin biosynthetic genes
JOURNAL Patent: WO 0151639-A 176 19-JUL-2001;
Schering Corporation (US)
FEATURES
SOURCE 1..1176
/organism="Micromonospora carbonacea"
/mol_type="unassigned DNA"
/db_xref="taxon:47853"
1..>1176
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC60078.1"
/db_xref="GI:15386337"
/translation="MMIKNGPVYRIRDLVRGKKVTIQTGYPTKTSKXNMMVQFRAEQ
LQGNALMPRGQITLADFYGEWMPSEYKTLKPTAVNSEGNIRNHLILTLDEL
DGOVTOQWVNDLEAGVPWPESTRGRKPLAAKTSINCHGLLHTICGAAIAKRLIN
PCSTMLPREPEKMFSLDPEIGRLITALPHWRPLVLLVATGLRGEALGRAGR
VDLLAARPLTVVEQLQELASTGELVFQSPKAKRRVTSFTTKVALLITPLIAGKKS
DEVVTFAPKGMVTRNFRIRIWKACEAGLPLGRIRHDLRHTHAALISAGRPISAI
RLGHSSIAVTDLLYGLHREVEDEGILAAIEAMAGVRAEDLEALDELDLVLADAA
"
CDS
ORIGIN
Query Match      99.7%; Score 1176; DB 6; Length 1176;
Best Local Similarity 100.0%; Pred. No. 1.2e-158;
Matches 1176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGTGGATCGAAGAACGGGCCGCTACCGCATTCGGGACCTGTTCCGGTAAAAG 60
DB 1 GTGTGGATCGAAGAACGGGCCGCTACCGCATTCGGGACCTGTTCCGGTAAAAG 60
```

Qy	61	GTCAACATTGACACCGGTTATCCGACGAAGACAGCGCCCAAGAATCGATGGTGCAAGTTC	120
Db	61	GTCAACATTGACACCGGTTATCCGACGAAGACAGCGCCCAAGAATCGATGGTGCAAGTTC	120
Qy	121	CGTGCAGGACAGTTGACGGCAACGGCTCATCCGCGCGCGCGTTCAGATTACCCCTCGCC	180
Db	121	CGTGCAGGACAGTTGACGGCAACGGCTCATCCGCGCGCGCGTTCAGATTACCCCTCGCC	180
Qy	181	GATTTTCGTGGGAGTGGTGGCCGAGCTACGAAAAAGACGCTGAAACCGACCGCGCTGAAC	240
Db	181	GATTTTCGTGGGAGTGGTGGCCGAGCTACGAAAAAGACGCTGAAACCGACCGCGCTGAAC	240
Qy	241	TCGAGGGGAACCGGATCCGAAACCACTCTGCCATACTCGGCCATCTCAACCCCTTGAC	300
Db	241	TCGAGGGGAACCGGATCCGAAACCACTCTGCCATACTCGGCCATCTCAACCCCTTGAC	300
Qy	301	GAGCTGACCGGACAGGTCAACCAAGCAGTGGGTCAACGACCTGGAGGCGCGCTCGGCCCG	360
Db	301	GAGCTGACCGGACAGGTCAACCAAGCAGTGGGTCAACGACCTGGAGGCGCGCTCGGCCCG	360
Qy	361	TGGCCGAGTCCACCGCGGCTGCTCGAAAGCGCTGGCAGCGAAGAGTACAGCAACTGC	420
Db	361	TGGCCGAGTCCACCGCGGCTGCTCGAAAGCGCTGGCAGCGAAGAGTACAGCAACTGC	420
Qy	421	CACGGCTGTCGACACAGTATCGGCGCGCGCGCATCGCGCGGAAAAGGATCAGGCTCAAC	480
Db	421	CACGGCTGTCGACACAGTATCGGCGCGCGCGCATCGCGCGGAAAAGGATCAGGCTCAAC	480
Qy	481	CCGTGCTCTTCGACGATGCTGCCCGCGCGAGCCGAAAGAGATGAAGTTCTTCGAGCGAC	540
Db	481	CCGTGCTCTTCGACGATGCTGCCCGCGCGAGCCGAAAGAGATGAAGTTCTTCGAGCGAC	540
Qy	541	CCGAGATCGTTCGGCTTATACGGCGCTTCGCGCGCACTGCGACCGCTCATGCTG	600
Db	541	CCGAGATCGTTCGGCTTATACGGCGCTTCGCGCGCACTGCGACCGCTCATGCTG	600
Qy	601	CTGTGCGACCGCTCTCAGGTGGGTGAGCGCATCGGCCTGCGCGCCCGCGGTCGAC	660
Db	601	CTGTGCGACCGCTCTCAGGTGGGTGAGCGCATCGGCCTGCGCGCCCGCGGTCGAC	660
Qy	661	CTGTTCGCGCGCGCGCCCGCTGACCGTTCGAGCAGTCCAGAGCTGGCCAGCAAG	720
Db	661	CTGTTCGCGCGCGCGCCCGCTGACCGTTCGAGCAGTCCAGAGCTGGCCAGCAAG	720
Qy	721	GGAGAGTCTGCTTTCAGTCCCGAAACCGCGAAGCGCGCGCGCGCGCGCTCAAC	780
Db	721	GGAGAGTCTGCTTTCAGTCCCGAAACCGCGAAGCGCGCGCGCGCGCTCAAC	780
Qy	781	ACGAAAGTCTGCTTTCAGTCCCGCACTATCCCGGAAAGAAAGTACGAGTCTG	840
Db	781	ACGAAAGTCTGCTTTCAGTCCCGCACTATCCCGGAAAGAAAGTACGAGTCTG	840
Qy	841	TTACCGCGCGGAAAGCGGATGTAAGGACCGCAATTCGCGCGGATCTGGGTCAAG	900
Db	841	TTACCGCGCGGAAAGCGGATGTAAGGACCGCAATTCGCGCGGATCTGGGTCAAG	900
Qy	901	CGGTGCGAGGAAGCGGCTTCGCGGTTACGATTCAGATTCGCGGACACTCAGCGC	960
Db	901	CGGTGCGAGGAAGCGGCTTCGCGGTTACGATTCAGATTCGCGGACACTCAGCGC	960
Qy	961	CGGATCTGATTTCTGCGCGGCTTCGCGGATCTCCCGCGCTTCGCGTCACTCG	1020
Db	961	CGGATCTGATTTCTGCGCGGCTTCGCGGATCTCCCGCGCTTCGCGTCACTCG	1020
Qy	1021	TCGATTCGCGGTCAGGATCTGTTACGGGCACTTCGCTGAGGAGTTCGAGGGATC	1080
Db	1021	TCGATTCGCGGTCAGGATCTGTTACGGGCACTTCGCTGAGGAGTTCGAGGGATC	1080
Qy	1081	CTCGCGGCGATCGAGGAGCGCATGCGCGCGCTCGGGCTGAGGACCTTGGAGCGGAACTC	1140
Db	1081	CTCGCGGCGATCGAGGAGCGCATGCGCGCGCTCGGGCTGAGGACCTTGGAGCGGAACTC	1140
Qy	1141	GACGAGGAGCTGACGAGCGTGTGGCCGACGACGA	1196

[illegible]

gene  
complement (2253. .2720)  
/gene="ml10004"  
CDS  
complement (2253. .2720)  
/gene="ml10004"  
/codon\_start=1  
/transl\_table=11  
/product="bacterioferritin"  
/protein\_id="BAB47684.1"  
/db\_xref="GI:14021071"  
/translation="MADLVGDVGLIGSPDNVKKHDKFKSKYNLTVDLADEERKVEIA  
YHLWVEKMYGRKYMVERATFLIDRDGRIARLWRKYRVKGAEDVLEAVRAL"  
2816. .6196  
/gene="mlr0006"  
2816. .6196  
/gene="mlr0006"  
/notes="unknown protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB47685.1"  
/db\_xref="GI:14021072"  
/translation="MDQEQPQHEKIRFRDEITDLGALPSCACRVPLPGQASIGRGFRI  
LRVFAGVAFVLLAAIVVYLVCAGSVGSEIRAEAEATIEKLAGVDVNVAVGPARI  
LSSSVLQVQSDVSKTSDDDKADAGVRFCMRLIPLLGEVRLTSARLSADIHV  
AAMPSDDMTADLRNEDGLVDPERLSDALFGNHRALDAVREDSNRIDLSNVEFLP  
ETCAIKLVADATVQSGPQAVRSEADIDGRKVTVAASREIDIAHRVTALTASV  
DUAADDTAAGAKGLGAVALKLTGAEGSANESERLTAUSSAGSVLDLGTGLLPADV  
DVATLVHGSNKIQVRLLLKTRGSFNFGSIGPKATGAAGEEPEPVRYDLTSDGST  
LAPSEPALFTFLAR IAGVYQTKSKLVAEQIAVRSAASEVLGTAAVAFVGGKAPG  
VNLALNHPDMPVSHVKLWPFSAENARLWDLNLFGRVVDANLQFQVPPDRLGAV  
PLSAEVEFRFOEGSRFTAGRIPIRDVAGVAFHGNVDVDSLSGTVVMPSGRTV  
AASNETVAKAANRPVICALDIDVEGAPVAELASVEPINAMRHVGFLEPDLGNTV  
TGHVRADIPLOGSVTSKLDWLVSLSLAKPEPQGTVDADGSIIVDEKAVIS  
AKASLNGIPAEIDLEPADDGPARKVALVLDKVRAMAPGLTPLLGGTVKVAID  
KSGDNQNVADLTNARLDIPWAGSKGAGVPAVTFVMTKAGDTTLLSDFNLGKTF  
SIDGIVLVNGLSKFSKVTLNRGDDVAVSVKRSKGVAVDINGNALDARSITKQF  
TSDVDTATVGTGTDNISVADVEQLTGFHDEQLSNLKLDSAGSRVNGLVKSIATSS  
GAATISNTDGRSLNVKSDAGAILFLNVRHBMGGSLMALTACGAGDGPMMKGV  
DTSNFVNEPKSLMNSTPAGDNRSNOAVKGNLDTSRVKEEFGYAEIEKSGYLK  
LANGVLRPRIGTTCGLIDONNMMDTGTFMNPYGLNRRIFGELFVIGALLNGNDR  
GLIGTVRGNAPKVDINPLSVIAPGIFRSIFEYR"  
complement (6203. .7456)  
/gene="ml10007"  
complement (6203. .7456)  
/gene="ml10007"  
/codon\_start=1  
/transl\_table=11  
/product="tyrosyl-tRNA synthetase"  
/protein\_id="BAB47686.1"  
/db\_xref="GI:14021073"  
/translation="MPAPKSDFLRTMSERGFTHQISDETGLDOLFAPKETTAYVGYDA  
TATSLHIGLISATWLYLQETGHPALMGGTSMDGPSFRDQSLTPEALATN  
LEGKRIEGRILRFQGDNDALMUNNADMLKLYVEILRDVGRHFSVNRMLTDSVK  
LELDREQSISLFENYMLIQGVDFVELARQNLQMGSGSQWGNINGVDLGRMKT  
POLYALTPTLLTSSGAKGKAGVWNLGDLFSFYDFQWWRNEDADVLEHFKIP  
TLPPLFEIARLAALGSGINEAKKILATETTAIVHGREANOAEETARKTFEAGALP  
TLPTVGDAAKALEAGIGISLLVLTAGLASSNGEARRHTQGGAVRINDQVSDDRRMT  
LQDLSPENVVKLGLKKHVLVRPA"  
7723. .8490  
/gene="mlr0009"  
7723. .8490  
/gene="mlr0009"  
/note="unknown protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB47687.1"  
/db\_xref="GI:14021074"  
/translation="MPHSTTBPRRMDPDLISVIPCRNEAANLPLLIDEIEAAMAGR  
HFELIIVNDGSTDETAALVITBOAALRSFPVRELHQRKSGAGSLVRSAGAAWARGIVA  
TIDTQGDNDPOYIPVLDALRQAGDFGAAGQRLKRDSSVKQLASRFANRLNAIL  
HDETRDTCGLKAVHETILRLKLPFDGTHRFVPAIVIQEYRVVHCDVDRSRHSGKS  
NYGIPDRGLQCALDLGGVWMLRRRRRRMRPKVEETKRV"  
8483. .8815

CDS  
/gene="mlr0010"  
8483. .8815  
/gene="mlr0010"  
/note="unknown protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB47688.1"  
/db\_xref="GI:14021075"  
/translation="MENVLOEGLTWLHVFIKQFDAMTLLIGFVAQGFTEFMEFVYOMLA  
SEKAKRSVPVAFWFFSLFGGGLLLIYAIQRODPVFIAGQANGMFIYIRNLWLIANER  
KAATKVD"  
8856. .10505  
/gene="mlr0011"  
8856. .10505  
/gene="mlr0011"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB47689.1"  
/db\_xref="GI:14021076"  
/translation="MALNRNYILFLFSLVMTLSGLASLPIDRDESRFVQATKOMAE  
SGTVDFRFOEASRYQKPIGYWLOSAAVTSLGKGAEPITWYRTISALGIAIAVLAI  
AWDNLFGANAGIAGVAAALIPATAFEGRDADKTDAWLLACCAAGALALIVLASR  
RNEPVAGHLEWIFKIAQGAALIKGPIAPILLSALTIAALFAFERDGRWLSKLRVGRGL  
LLVVITPLWLAALTWKSAGFIQQAQVGMKGVASGEESHGLPPGFYMLTYSLFWW  
FGLIATVGLQALNRRDVRFLRFLAWYIPFWLVEFLIPTKLPHVLYPAYPGMALL  
IGWLLTIQPEDANASLRWQWLMWSTAFGLVYVSLGAAVISCAPILYTHSFWMSI  
PAAALQGTDFAFSRQQLVPLRIGATACAGITALLFGVIAPSKPLTWLSPAIKO  
AVLANKCDMTVLASARFOEPSLVLVGTVTTLTVDGVAQHLLTDPACALGLAPIED  
EOKLNGMSYQGVSKVRAEIDGLNYSYGKLSGLRVAP"  
10632. .11366  
/gene="mlr0012"  
10632. .11366  
/gene="mlr0012"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB47690.1"  
/db\_xref="GI:14021077"  
/translation="MRIVGRFAARPARYPNIAWPVWGLVLLLTAVAFVRLDTPAGM  
VHQSGAGLAELITQFALGGWYLIPLSALLVAANLTDRGLSRRSLMNVNMTCLAF  
LVLSAVGLGLLVNVLKYAIGRAPLYFQDFGLVALHPFAFDARFAGFPGSHATTMGA  
VFGVLLLPFRWHIALVVTACIATRVPVGAHVPSDVTAGVGLGCAAFALACGLVFRAR  
LGFIIRATPSGLPVKASPELIAPQK"  
complement (11379. .12029)  
/gene="ml10013"  
complement (11379. .12029)  
/gene="ml10013"  
/codon\_start=1  
/transl\_table=11  
/product="transcriptional regulator"  
/protein\_id="BAB47691.1"  
/db\_xref="GI:14021078"  
/translation="MYEPHPQETRPDLVHLIRAHPLGMLISNGPDGPVANAIPFLI  
DAEGLNGRLRAHLAKANPHWRIADNPASPLVIFQGDADAVTTPSWYETKRTGKV  
PTWYAIQVVRGTAKVIDQEWLAQIADLTGQEGARQAPNAVTDAPPPFIOQIKG  
IIGLEIETSHGKWKVSONRPAVDRAAGVAQGLESEVADASDVARLVRSYGGLDN"  
complement (12051. .12737)  
/gene="ml10014"  
complement (12051. .12737)  
Query Match 5.6%; Score 66.4; DB 1; Length 347660;  
Best Local Similarity 44.9%; Pred. No. 0.79;  
Matches 329; Conservative 0; Mismatches 376; Indels 27; Gaps 1;  
Qy 348 CGCGCTCGGCGCTGGCCGAGTCCACGCGGCTGCTCGGAGCCGCTGCACCGAGAC 407  
Dy 348935 CGTCGACACCGGACGAGAGGATCAGCGCGGCGCTTCGCCCTCGGCTCGCAC  
Qy 408 GATCAGCAACTGCCAGCGCTGCTGCACACGATCTCGGCGCGGCGGATCGCGCGCAACG 467  
Dy 346895 AGTCCACCACTACCGAGCGGTGCTCGTGAAGGCTTGGGCCAGCGCTGACCTGGGACG 346894

RESULT 5	AP005036/c	300800 bp	DNA	linear	BCT 10-MAY-2003
LOCUS	AP005036/c				
DEFINITION	Streptomyces avermitilis genomic DNA, complete genome, section 16/30.				
ACCESSION	AP005036	BA000030			
VERSION	AP005036.1	GI:29607295			
KEYWORDS	Streptomyces avermitilis MA-4680				
SOURCE	Streptomyces avermitilis MA-4680				
ORGANISM	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.				
REFERENCE	1	Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M.			
AUTHORS	Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites				
TITLE	Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)				
JOURNAL	21477403				
MEDLINE	11572948				
PUBMED					
REFERENCE	2	Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M. and Omura, S.			
AUTHORS	Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis				
TITLE	Nat. Biotechnol. 21 (5), 526-531 (2003)				
JOURNAL					

FEATURES  
source

RTTUMRFSYHDFDQ

1..300800  
/organism="Streptomyces avermectinis MA-4680"  
/mol\_type="Genomic DNA"  
/strain="MA-4680"  
/db\_xref="taxon:227882"  
/note="This strain is also named as strain: ATCC 31267,  
NCIMB 12804 or NRRL 8165.-synonym: Streptomyces  
avermectinius"  
complement(84..11144)

```

complement(84, .11144)
$names<-SAV3642
nms2 gene cluster"
/codon_start=1
/transl_table=1
product="putative non-ribosomal peptide synthetase"
protein_id="BAC71354.1"
/db_xref="GI:29607296"
/transl_table="MEKPSVSGTSDADRALAEELRRRLAGRRGTGATRRPDPADRRIR
LPLSVGQQQMFLSLEPDSERYVLVLRLGALDEALDEAKWHLRHHETRLRL
ADAGGEPVQVIDAPRLPAGSGSGAADDALVRLAEADAPLPDGLDMPVRAELIR
DDEHVLVWFHIIADAMSTGVSGLGSDALVTSAPLPIVQVADVYAAWQRAELT
DEHVLHKKYGLADLAPLETDRIRAVTGRFVADGSGEAVDPAPLGRTRLEAA
TFFMLFHWAGSLARLETGRADIVTGVSGRGLPGLVGLIGVNTLVNRGLSS
TFTLLATRAVTDAYDQHDQVAVQADALGSLRTEFLQVATVHGROSS
TFFPVGVAPEFAGARVAKEDQLQIREADGSGFGLHRLRYATLDFRATVRFAGHL
TGLTAAATPGARLSANVIREKEVALGLCAPEAVPVRVHEVFEQVATPGVAA
PGSTGLVYAEILNAPNRV1GAHILAKGVGPETLVGLSERGIEI1PALGGLKSGA
PDPNPANADRTAIVGDAQAAVVVTAHEHLLFDEVLLELTDAPLAAFPDSDP
SGSPENLYITLTSYGRPKGVKALTHNVRLLEGRNEHYAFTDVPVLFHSAVD
SVWENGMALHGGRLVTVSPFEDLLVRENTVNTQPAFASLVAAD
SURVAFAGEKLETSELSPADWRGLDRLANVMGIVTETVHTTTRLEATRPV
GNAVHAPDRLRVLLDQALQVFPVGEIHVGPGCVAGYLNRLPHEATRPVDP
GPAGRLYKSGDLARLPDGLSLEGLIDDOVKIRFIBLGE1ETALAALPQVR
VLVBREDPDKRLVAYTTPAADOAPGCDLRSHLAARLPMVYPAFVALDALP
GKLDKRALPQCDALGACHVATPVAEBRI1ABWVALGLDVRGVGDEPFLGQ
TRAVTLVGLRLA1GYDLAAVROFVRYTVAEVLVTCGRPADEADRTBFELTY
AVWAHAEALRTSDVTGYSVQLQVHLHTTAENVGARDLTSLEELL3SLRETA
SILFDATPLRLHAGWAGDSGKWI1NTECHAILDGWHSLLMEVIEBGRVRI
QFSEAPVGFVREDFAEALDVLDSAGTTRAYQSVQDGYARLTVAPWQDPODT
RYRVYVFPDLEDRLEATLKAGASLKSVLHAHLTKLSLMEEDRFSGVLGNAR
LGADRYGVGHMLNT1FAYDRTAAVNDLAAAFARELELWPHRT1PMPV1ORLEA
RT1DWFYSVHDPOVDVRAQVDYAS1DSD1PFEFLGPGVSARVGHVLVITASPKALR

```

DLAAMLRGVLEMAADPEGDAQASFLPAGERERQVLEWNTAYEABTASVLRREFEQ  
AARTPDRAVAGHQDGLSYGELDVANQALHRLGLGVGAERSVLRDRGPDLTDL  
LAWKGAAPVDPVDPADRTAMEELSGAPVLTETIPVRLAISLQDDESHVAPGPVERVA  
RLAYVLTSGSTGTQGVETVTHRGHANVAMARELASQDQCAPLFSVVAQDLVVPN  
LWAPLVGTQVHTVQSDQMDADLGEHLVAGSPYSFKLTPGHLVDVLAQOALPEQAAAL  
APVLVVAGEAFTATLQERWALAPDRLINEXGPTASVGTTHVEPRDAEDVLPJIG  
RPLPNRVVLDLALQVPVGVGTGELVVGSTGARGYAGRPDLTARFLPDYGTGTPG  
ARYRTGDLVRQPDGNVGFPLGRVDQVKIRGRVVELGEIQAVLTAHPGIRGVTVV  
DDELAGYTPAGAEVREHLADRLPDYMPVATLTDALPLNANGIKDRKALPAPDA  
AADDDGVPARTTEERIAAAMQVIGLERVEGDDFFEIGHSIRAVALVAGLRASAG  
FDIGRDFVEHRTVAALGHELTGRPARTVETVPARFALVPAEDARLPEDAVDAPL  
TQVLQGLWLEMLADDTKHAHVNSFRITEAPPSADALRGAAAVVAGRHEMLRTSFH  
LDGFSLOPVVRAHAALPVAHDLTALDAGRRRAVEEPIABERADLPLERAPILLEM  
TAHVEDGAWRLGLTCHALTEGWSHRALLMELLDYRLRGQDQSEPAAPAVIAD  
FVAIEVRSSEHRYARDVDTGCFARFALPAWAGDRAPRETFKVAVELPFPFQGL  
RALAARSLKSVLLAAHLKALSVLEEDRFHSGLVFGFAPEACADRVYGVYLYNTL  
RPAADRRLTWELVREYFAHEAGNWEYRYPMPAIQDAGVERLLDVRFYQDFDH  
VDFRVDSSGEGATEFSLAVSANDVSTYMTTHALSANADRLTALYAGVLTAM  
TQPDAGDAVPLADDELKQLVWNTAYEABTASVLRREFEQARTPRAPAVTLGE  
RLTYAELDEHANLALHARGVGAESVAVQLORGPVLIJALLAVMKAGAYVPVD  
PSYPAERVASIVGTSAAVANVSAAYADRFAGVSLLLDDEDVSALPSAPARADLDQ  
LAIVTISGSTRGKGVFPHRGLAGHVGWAARELASRGTGGAPLFSVAFDLIIVPNL  
WAPLVGTQVFAVQDTEPAELGHRLLAGAPYSFVKLTPEGLDLAEQUTTEQAQELA  
EVMVAGEALPGATADRLSLDILPGRLINEXGPTASVGTTHVEPRDAEDVLPJIGR  
PLPNRVVLDLALQVPVGVGTGELVVGSTGARGYAGRPDLTARFLPDYAAEPGA  
RLYRTGDLVRLPDGNEVFLGRVDQVKIRGRVVELGEIQAVLTAHPGIRGVTVV  
PVPGRRLVGYMTPAAGACAPPELAACHAVRLPDYMPVSAFVALEALPLNANGVDR  
GVLPVPGRCALSGTHVPRPTKALGAINSEVLVEEIGVHPRFEDLGGHSLMI  
KYLAAARAGLAVSTRMYOHDTLIDLAADVETDAGARAGSQSADAESATGAAA  
ETQASQTPASAEAGSPFAEASAVGAVSAGGARASGAAGSSGEAVPAVAAVGS  
SEBAVFTAAAAPVPTPAGVDAAAPTAVVSAELLAALEQAAGGPFGRAGSGVVP  
EAAALGGAAREAPVGRGDRDPLAELTAAMAEHRVPGVALAVRDGAVARLGHGT  
LEAGGSDPVYRTLFQVGSISKHLTAFVRLVHEGLIDLDADIDGYTTSMLADPA  
APCTVTRVLLGHRAGLARHRSVGRPGQLPTLLDLLEGPRVPTTPVRRLPGEV  
FKKSHVYVWLQVLEDTVGEPELIMRLRLDPLGLADTSFDQHPRTAGLPLVALLH  
DAHGTPLSGRERLAAGLWNTAGDAARFVLAVRDALLGAPALVQALARELHA  
GHTGTPYGLTIVDDTDLDFEGHGEPSPGYNNMAISRMHGTTGFVALTNADSGKAVV  
KYLTAELGRDRTAFSGRLAADRREDPAGALPAVLAVPVVTEERA"  
complement (11156..16285)  
/notes="SAV3643  
nrps2 gene cluster,  
dihydroxy-benzoyl-glycine biosynthesis"  
/codon\_start=1  
/transl\_table=1  
/product="putative non-ribosomal peptide synthetase"  
/protein\_id="BAC71355.1"  
/db\_xref="GI:29607297"  
/translation="MNAELLSERFAEQFRTHAAAVDGAFTVYIELDRASLRAR  
HLRDGAGPETLVGSLRGVLDLIIVALLAVRAGAYVPLDPAQPPARSLDLARAGA  
RLVAGAGPVRDAGARVGPPEIPDRSSDPAGLPAAADPANAAVAVFTSGSTRP  
KAVVTHAGIANRIGMTVQVHALGAADRVLQXTTIGFDAAGWEIIFAPLVGGTVVLA  
AGARDPAALLRAVDHGVTVLVVPSVLRLLVEGDSWGGSLRLLSAGEALHAEL  
VARLRETDQDLVWNTYGTETCSIDITAQVLDVPLTAGVPVIGRPLPQMRVILVGN  
GVPPVGVPEGLVAGGVGARGVAGRPDLTADRFPVDPVATEPGARLYRTGDQVWRS  
DRTLEYGLSHQVKNVRIEPAVEEALAAHPAVTGAIVTPYADGGKRLAAYLT  
VSGEADPAGRLGLAERLDPDSHVSFLHTLDPLUTANGKVDRAALPSPAEIAAGQP  
TYTARDAAEGLVAVWADLLGLDRVGLDDFFALGGTSLQTLRAALRAASGEKVS  
LREGLAATVTAQAQLIGTTPGAADGTSAGVGTGTGAGHEGTGCGAVHGGTGD  
VDDQIPVPTGDLPLSSGQRLWFLDRHPGSPFEWVAPLFLRPLSGLDAGTVGRA  
DLAALAEHPLETRVYDQGGEPQIIDAPGAGTVELRVEDTARDVAAALFGAOPERG  
DLAAGPLRALLARVAGEDHVLMTMHITCDGWSTVILEREILRELCADRAGAPDL  
PELQYADYGSQCARLTDAVEGELTHRAALDGLAPLDLPADHPAPVDRDRAV  
VTPVPAPELAKVLAELGRAHGAFTFMILLTAYTAVRHITGOWDPVGTVPAGKTRPE  
TENNYGFLNLSVRCGLTADUGFTALURVDRDAALVHQBELFVELSELOPERD  
LSRTPLYQVAFDLHSEGVTSVYTHSDLAAPFAWQVAKTDLISFMFRTPDGSLDVL  
EYATSLFERGTAEHTGHFLTLLEQVAEQPETPLGALLSADSHRLLVWNTDAED  
TDRVTVEFEQASAPERTALTGETVYTGELDAANRLAQLHRLVRCVGRASRA  
VLLDRGPFLVTTALLAVKAGAVVVDPSYPAERIAAMCDTAGVRTAVTTSAYARFT  
SSGTRLLDLDADILAGCPAAAPARTDRPRLAYTIETSGSTGTGPKVETVHOGLAN  
HYMAARELVLQDGGGALLSSAAFDLVVPLNWLAPLVGTQKRLFLPLQDIMSGLKRL  
ADAQPFVFLKGTGHDILALQUTPAQAGALAPLVVAGEAFTATLREWRALAPDTR  
LINEYGPTEASVGTTVTPIEHADADVPIGRPLENRMVYVLDPALQVPVGVAGELY  
VGGTGVARGYANRPDLTADRFPLDPYGTGPGARMYRTGDLVRLPDGNVAFVLRVDDQ

VKIRGYRVELGEIQAVLTAHPAVDRAVVTVRHPENGEPPTLAAHVVPADAAAPLPLDAA  
HCAARLPQYMIPTATFALDTPVNAVGNVDRSALPAPORTADDESHVAPGPVERVA  
BIWTELLQVQAGAHNPFHIGNSLILRLISHLQQBFEIDFAVRTVFEQPTVARIIA  
TVBERVTAQIAALSDADLLDDAARTDITHTLNNALKHEQA"  
complement (16329..16595)  
/note="SAV3644  
nrps2 gene cluster"  
/codon\_start=1  
/transl\_table=1  
/product="putative MbH-like protein"  
/protein\_id="BAC71356.1"  
/db\_xref="GI:29607298"  
/translation="MTAEQTAQAQADPAQAQAQCEGVPORVFNDEEQVSLMPAGR  
DLPAGWAEAGTEGSRRECLARIAEVDWTDMLRPLSLRRHMHSSAA"  
complement (16629..17633)  
/gene="arcB2"  
complement (16629..17633)  
/gene="arcB2"  
/note="SAV3645  
nrps2 gene cluster"  
/codon\_start=1  
/transl\_table=1  
/product="putative ornithine cyclodeaminase"  
/protein\_id="BAC71357.1"  
/db\_xref="GI:29607299"  
/translation="MLILRADVTDVLSGRETEILDLVAETRYLRHDSQTSPLHSTFL  
RFPBEHRSRRIIGLPAYRGGERVAGKMWISSPFGVNAVAGTDRASAADVLSLNGH  
PVAFVEGAVISAKETAASALAARELTADNPFAAALLIGCVINLEILRFLAALPLD

Query Match	5.6%	Score 65.8	DB 1	Length 300800
Best Local Similarity	48.9%	Pred. No. 0.99		
Matches 280	Conservative 0	Mismatches 272	Indels 21	Gaps 3
QY 393	GCTGGCAGCAAGACGATCACCACTGCCACGGCTCTGCACACGATCTCCGGCGCGGC	452		
DB 115551	GGTGGGAGACCCACGGTGTCAAGGCGTACCACTGCTGGCGCCCTGATGAACACCGC	115492		
QY 453	GATCGCGCGAAAACGGATCAGCTCAACCCGCTGCTTTCAGACGATGCTGCCCGCGCGA	512		
DB 115491	GGTAGATGACAGCTGATCCCGCAACCCGTGCGGATCAAGCGCGGACAGTACGA	115432		
QY 513	CCGAAAGAGATGAAGTTCCTGAGCGACCCCGAGATCGTGGCTTATCAGCGCGCTCC	572		
DB 115431	CGTGCTGAGCGGCTGTTTTGTCGGTGGCGGAAGTCTTTGGCGCTCGCTATGCCATTGC	115372		
QY 573	GCCGACTGGCGACCGCTCGTCATGCTGCTGGTGGCGACCGGTCTGAGGTGGGTGAGGC	632		
DB 115371	ACCGGCTACCGGCTGCTGCTCTCTGCTGGTGGCGTTTCCAGCTCTCCGCTTCGGCGAGCT	115312		
QY 633	GATCGGCTGCGCGCCCGCGGTCGACCTGCTCGCGCGCGCGCCCGGCTGACCGCTCGT	692		
DB 115311	GGCGCGCTGCGTCGCAAGGACATCGATCTGAACGGT-----CGCTTGTGTGGT	115261		
QY 693	CGACAGCTCCAGAGCTGGCCAGCAGGAGAGCTGCTTCCAGTCGCGGAAGACCGC	752		
DB 115260	CCGCGGGCACAGCCCGAGTTGCAGAACGGCGCGCTCTTCAGAACGGCGCCCAAGTCGC	115201		
QY 753	GAAGCGCGCGCGACCGCTCAGTTTCCACCAAGAAAGTCGCTTACTGCTTACGCCACCTCAT	812		
DB 115200	CGCGGGCTTCGAAACCGTCTTCCCGCTGAGCTGCTCGACGAGATCAGCCACACCT	115141		
QY 813	-----CGCGGAAGAAAGTACGAGGTCTGTTTCCCGCGCGGAAAGCGCGGAT	863		
DB 115140	GAAGCACTACCGCGCCGATGTCAGGAGGGCCACCTTCTGTTGGTGCACAGGTGGGCA	115081		
QY 864	GGAAGCAGCGCAATTTCCGCGGATCTGGGTCAAGCGCTGCGAGGAAGCGCGGCTTC	923		
DB 115080	GCITCGCGGAGCACTTCCGGGATGATCGATCAAGCCAGGACGACGCGGCGCTCTC	115021		
QY 924	GGGT----TAGGATTCACGATCTGCGGCAC	953		
DB 115020	GCCCAAGCTCACTTTTCACGACCTGAGGCATAC	114988		

```

RESULT 6
AF429315/c
LOCUS Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
DEFINITION AF429315
ACCESSION AF429315
VERSION AF429315.1 GI:17646244
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 125020)
AUTHORS Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
Porter,N.T., Ross,C.A. and Margolis,R.L.
TITLE A repeat expansion in the gene encoding junctophilin-3 is
associated with Huntington disease-like 2
JOURNAL Nat. Genet. 29 (4), 377-378 (2001)
MEDLINE 21583737
PUBMED 11694876
REFERENCE 2 (bases 1 to 125020)
AUTHORS Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
Direct Submission
TITLE Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
LOCATION/Qualifiers
FEATURES
source
1..125020
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q24.3; between D16S520 and WI-12410"
/notes="isolated from a patient with Huntington's
Disease-like 2 (HDL2)"
complement(35581..35746)
/rpt_type=tandem
/rpt_unit="ctg"
complement(<36507..>36887)
/gene="JPH3"
/notes="synonym: JP3"
complement(<36507..>36887)
/gene="JPH3"
/product="junctophilin 3"
complement(<36507..>36887)
/gene="JPH3"
/notes="component of the junctional complex between plasma
membrane and endoplasmic reticulum"
/codon_start=1
/product="junctophilin 3"
/protein_id="AAL40941.1"
/db_xref="GI:17646245"
/translators="M5SCGRFNFDDGSGYCGHEDGKAHCHGVCTGPKGGVGTGWS
HGFEVLGVYTPSGNTYGTWAOQKRGIGLESKRWVINGEWHGFKRYGVRECA
NGAKYEGTWNGLDQGYGTETSDG"

Query Match 5.6%; Score 65.6; DB 9; Length 125020;
Best Local Similarity 11.3%; Pred. No. 1.3;
Matches 100; Conservative 390; Mismatches 381; Indels 12; Gaps 3;

QY 254 GGATCCGCAACACCTCTCTCCCATCTCGGAGGCGGCGCTCGGAGGCGGCGGAGTCCA 313
D 17766 GKGKSSKKWKSSTRRSSAKSCSYMWGSMCMSCMSMAKSYCYMYRMSY 17707
QY 314 AGGTACCCAGCAGTGGTCAACACCTCGGAGGCGGCGCTCGGAGGCGGCGGAGTCCA 373
D 17706 SYRYKVCSCMGSTSYSCCCKWWSGSCYKCKYKYSWSSYSTSMGYSTSCCKYK 17647
QY 374 CGCGGGGTCTGCGGAAGCCGTGGCAGCAGGATCAGCAACTGCACGGCTGCTGC 433
D 17646 CSWSSMYKCKTSKYRRKRSYVYVGGGKRAKKYKACGRRRRMSYMKCAKWMSSYCCWS 17587

```

```

QY 434 ACAGATCTCGGCGGCGGCGATCGGCGGAAACGGATCAGGCTCAACCCGCTCTTCGA 493
D 17586 YCMYTYYSKCTKYKSTCYKRGYGYGSKTCYAGGKSRSMYCMRKRSSKSSWSMAR 17527
QY 494 CGATGCTGCGGCGGCGGAGCGGAGAGATGAAGTTCTCTGAGCAGCAGCCGAGATCGGTC 553
D 17526 SSWCMGWA---CYRKRKSAGWAGRSSKGRSTGKVRACSKKTGSGYSTGRSMKK 17471
QY 554 GGCTTATACAGGCGCTTCGCGGCACTGGCGACCGCTGCTCATCTGCTGCTGCGGACCG 613
D 17470 KGYSKSRGKMKKKKTCYCMWYKYKRTSMCWYMYKSRKRCMKKKGCTGYRG 17411
QY 614 GTCGAGTGGGAGGCGATCGGCTCGCGCGCGCGGCTCGACCTCTCTCGCGCGC 673
D 17410 MSKKSGYKMRGSRGSSYSTSCWKSCHGYSWCKMYMYKRRSRMRSMSSKGRGY 17351
QY 674 GCGCGCGGCTGACCGCTGCTGAGCAGCTCCAGAGCTGGCAGCAGGAGAGCTCTCT 733
D 17350 AGRGYS--MWSTKRKRSKYSYKYGKRGKMGKMGKRGSKYKSSMKQKRSWSK 17292
QY 734 TCCAGTCCGAGACCGGAGCGGCGGCGGCGGCTCAGTTTCAACCAAGTCTGCTC 793
D 17291 CYSTKSYKSRGSRKSGWRSTSKAKSSMRMAGSKCTYGSYSNNRNNRMKTCGNM 17232
QY 794 TACTGCTTACGCCACTCATCGCC-----GGAAAGAAAGTACGAGGTCGTGTTCA 846
D 17231 YERSPMNNGNAAAGCTTCCCCANTNGGGGAAAGGCGGASRASCYKGRMSSKCY 17172
QY 847 CGCGCGAAAGCGGAGTGGTAAGACCGCAATTCGCGGCGGATCTGGTCAAGCGGTC 906
D 17171 RSGTTRRCMKSKCRRSRGSKSMGTGRSGKTSYSAKSGRGYCYCWGMRGRKRCMSS 17112
QY 907 GAGGAAGCGGCGCTTCGCGGCTTACGATTCACCATTCAGGCTCGCGGACACTCACGCGGCGATC 966
D 17111 RMKNYKSYRRKRWMTCKMYCYMYMYCRSMCMCKSCCGCYSMGMSYSGYK 17052
QY 967 CTGATTTCTCGCGGCGCTGCTGCGGATCTCCCGCGGCTCGCTCTCTCTCTCTCT 1026
D 17051 SWGKSYMRSYYSKRSKTSKAWRSKRGMTGTRGGRSYKGGGSGYKGGSGWGGKG 16992
QY 1027 CGGTCACGAGTCTGCTACGCGCACTCGCTGAGGAGTTCGAGGAGGATCTCTCGCG 1086
D 16991 KSTRGSSAGKSKSYKMSKCCARYKMSYKCYSRMCMCMYCYCWGCKKSSCWSSMRK 16932
QY 1087 CGGATCGAGGAGGCGATGCGCGGCTCGCGGCTCGGAGGAGGCTGG 1129
D 16931 KSKGGSWAKGMVMDKGSVSTDKSDBSRBSKVKSKWSMR 16889

RESULT 7
AF429315
LOCUS Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
DEFINITION AF429315
ACCESSION AF429315.1 GI:17646244
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 125020)
AUTHORS Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
Porter,N.T., Ross,C.A. and Margolis,R.L.
TITLE A repeat expansion in the gene encoding junctophilin-3 is
associated with Huntington disease-like 2
JOURNAL Nat. Genet. 29 (4), 377-378 (2001)
MEDLINE 21583737
PUBMED 11694876
REFERENCE 2 (bases 1 to 125020)
AUTHORS Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
Direct Submission
TITLE Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
LOCATION/Qualifiers
FEATURES
source
1..125020
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q24.3; between D16S520 and WI-12410"
/notes="isolated from a patient with Huntington's
Disease-like 2 (HDL2)"
complement(35581..35746)
/rpt_type=tandem
/rpt_unit="ctg"
complement(<36507..>36887)
/gene="JPH3"
/notes="synonym: JP3"
complement(<36507..>36887)
/gene="JPH3"
/product="junctophilin 3"
complement(<36507..>36887)
/gene="JPH3"
/notes="component of the junctional complex between plasma
membrane and endoplasmic reticulum"
/codon_start=1
/product="junctophilin 3"
/protein_id="AAL40941.1"
/db_xref="GI:17646245"
/translators="M5SCGRFNFDDGSGYCGHEDGKAHCHGVCTGPKGGVGTGWS
HGFEVLGVYTPSGNTYGTWAOQKRGIGLESKRWVINGEWHGFKRYGVRECA
NGAKYEGTWNGLDQGYGTETSDG"

```

```

Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
Location/Qualifiers
1. 125020
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q24.3; between D16S520 and WI-12410"
/note="isolated from a patient with Huntington's
Disease-Like 2 (HDL2)"
complement(35581..35746)
/rpt_type=tandem
/rpt_unit="ctg"
complement(36507..36887)
/genes="JPH3"
/note="synonym: JP3"
complement(36507..36887)
/genes="JPH3"
/product="junctophilin 3"
complement(36507..36887)
/genes="JPH3"
/note="component of the junctional complex between plasma
membrane and endoplasmic reticulum"
/codon_start=1
/product="junctophilin 3"
/protein_id="AAL40941.1"
/db_xref="GI:17648245"
/translation="MSSGGRFNDDGGSCGGWEDKAGHGVCTGPKQGGHYTGSGWS
HGFEVLGVTPSGNTYQGTWAGQKHGIGLESKGVYKGEWTFGFRGYGVRECA
NGAKYEGTWSNGLQDGYGTETYSYG"

ORIGIN
Query Match 5.3%; Score 62.4; DB 9; Length 125020;
Best Local Similarity 12.2%; Pred. No. 3.7; Mismatches 392; Indels 9; Gaps 4;
Matches 109; Conservative 387;

QY 126 GGACGAGTTCGAGGCAACCGCTCATCGCGCGCGGTGAGATTACCTCGCGCATTT 185
D 16948 GRRRSKGMWYSGMRSGRSKMYTGGSKMRSSMCTSCYASMCWCWCSCCVRSCC 17007

QY 186 CGTGGGGAGTGTGCGGAGCTACGAAAGACGTGAACGACGCGCGTGAATCGGA 245
D 17008 MCRSYCCWRYCACKCYSSYVMTSASYNRRSYWKRSMKWSRSGRSKCKSRGC 17067

QY 246 GGGCAACCGGATCCGCAACCACTCTCGCCATATCGGCCATCTCACCCITGACGAGT 305
D 17068 GSGMGKGGKSYGRKTKSRGKMGKAWYMYRSMKMYSKGYCMYCWGRR 17127

QY 306 GGACGGGAGGTACCCACAGTGGGTCAACGACTGGAGGCGCGCGTGGCCCGTGGCC 365
D 17128 GCYCSMTSRSAMCCSYCAKCKSMCYCYGMSMKYGYACSYRGSMSKTCMRGTSY 17187

QY 366 GGAGTCCACGCGGGTGTGCGGAAGCGGTGGCAGGAAGATGATCAGCACTGCCACG 425
D 17188 TSGCGCTTTTCCCCNANTGGGAGAGTTTNNCKITYSYRRNGCAGC-KYNNYN 17246

QY 426 CTGTGTGACAGATCTCGGCGCGCGGATCGCGCGAAGCGATCAGCTCAACCGGTG 485
D 17247 NSWRSSCRAGMCTKYCKSMNTMASYCMWMSMYCSMRSMASGWSWSSYMMKSSWR 17306

QY 486 CTCTTCGACGATGCTGCCCGCGGAGCGGAAAGATGAGTTCCTGAGGACCCGGA 545
D 17307 MGYCKCCKWCMYCMMSRMSRGNYSYMYASWSSSRGVCYCTRCYCMWSKSKCYKYY 17366

QY 546 GATCGTGGCTTATCAGCGCGCTTCCGCGCACTGGGACCGCTCGTCACTGCTCGT 605
D 17367 MMRSKRMKGMKWSRWGSMWGSASRSYCYKYSKMRCSMSSKCYR--CAGNNMKGY 17424

QY 606 GCGGACCGTCTGAGTGGGTGAGCGATCGCGCTGCGCGCGCGCGGTGACCTGCT 665
D 17425 MYMRCSWKRMRWGWKSAFYRRYWKRGAMMCMKCYSRMSRCMMWKKYSCASCRSA 17484

QY 666 CGCCGCGCGCGCGCTGACCGTCTGTCGAGCAGCTCCAGGAGTGGCCAGCAGGAGA 725

FEATURES
Db 17485 MWSGYMKCASYNWCMSSYWCCTCWCTSYMSYVRCCTCWCKGSSSYTKSSWSMSSSYKGR 17544
QY 726 GCTGCTCTTCAGTCCGCGAGAGACGCGGAAGCGCGCGCGCGTCACTTCACACGAA 785
Db 17545 R---KSYSMCTTSGAMSCWRCCYMRGASNRAGSMRRRAKGRSGRRKMMWTGGMW 17601
QY 786 AGTCGCTCTACTGTTAGCCACTCATCGCGGAAAGAAAGTACGAGGTCGTGTTTCC 845
Db 17602 RSKYVYCTGRRMMWMTYMCWCWRRSYMYRSMAMGMRKSSWSGMRMMGSSARRCKSA 17661
QY 846 CGCCCGGAAAGCGGATGTTAAGACGCGCAATTCGCGCGGATCTGGGTCAAGCGTG 905
Db 17662 SRSSWCSRRMKMRGSCWSSKMGWGRSRRSASSCKGSRGMRSSKSSKRYKRGGRKKR 17721
QY 906 CGAGGAAGCCGGGTTCCG---GGCTTACGCAATTCAGCATCTGCGGCACACTCACGGCG 962
Db 17722 SMTKSGSKGSKCKWKRSGSMTSSCYYSASSCMMWSSKSCMCCMMKRCACCTSSM 17781
QY 963 GATCTGATTTCTCGCGGCGTCTGCTGTCGGCGATCTCCGCGCGCTCGGTCACTC 1019
Db 17782 SCTSMYRCWGMGMSYSTCGTCTCKCTGYKKSRTWYMMYSWTSKTYMWRACCMYC 17838

RESULT 8
LOCUS A81401 2404 bp DNA linear PAT 21-JAN-2000
DEFINITION Sequence 1 from Patent WO9907861.
ACCESSION A81401
VERSION A81401.1 GI:6731721
KEYWORDS
SOURCE Mycobacterium phage M66
ORGANISM Mycobacterium phage M66
REFERENCE 1 (bases 1 to 2404)
AUTHORS Da,C.G., Freitas,V.A. and Trans-Complementation,U.A.
JOURNAL Patent: WO 9907861-A 1 18-FEB-1999;
DA COSTA GARCIA MIGUEL ANGELO (PT); FREITAS VIEIRA ALCINO (PT)
FEATURES
source
1. 2404
/organism="Mycobacterium phage M66"
/mol_type="unassigned DNA"
/db_xref="taxon:65388"
697..1815
/codon_start=1
/transl_table=11
/product="INTEGRASE"
/protein_id="CAB69190.1"
/db_xref="GI:6731722"
/translation="MASVRERVRKDGTTAYLVSYRTFGSGSAOQALTFDNRKAADAPA
AAYDAHGAASALEMHWGINPAPRGTKSELTVAEWIRHHIHLTGVEQYTDKXBOYLAN
DIBNLGDIPLSKLSBEDIARWYKVMETGDRDNGHAPKTLNKKYGLSGALNAAVP
RYLSTNPASGRRLPRGNAEDDEIRMLTHAEFDRLRDVATPHWKLKMWQFVWSTGLRWG
EFVYVNTDGGPVRYPGFLRRVMPAVEKAGLVRPTPHDLRHTYASWQLTGGTGTIV
SRQLGHESIQTIVDTVTDVDRTSRRAAEFMDGLLGF"
```

Db 1110 GCACGCCCGAAACCCCTCCGCAACAAATACGG---GTTCTTATCGGGGGCACTGAACGC 1166  
 QY 437 CGATCTCGCGCGCGGATCGGGCGAAACCGATCAGGCTCAACCGTCTCTTCGACGA 496  
 Db 1167 CGCGGTCCCGCGATATGTTGCACCAACCTTGGTGGCGCGCGCTGCCCGTGGGNA 1226  
 QY 497 TGTGTCGCCCGCGCGAGAGAGATGAAGTTCCTGAGACACCGCGAGATCGGTCCGC 556  
 Db 1227 CGCTG-----AGGACGACGACGATCCGATGCTCACCCACGCGGAGTTCGACCGC 1279  
 QY 557 TTATCAGCGCTTCGCGCGATCGCGACCTCGTCTGCTGCTGTTGGCGACCGGTC 616  
 Db 1280 TCCGCGACCGGTGACACCTCACTGGAAGCTATGTTGATGTTGATGTTGATGTTG 1339  
 QY 617 TGAAGTGGGTGAGGCGATCGGCTCGCGCGCGCGCGGTTCGACCTGCTCGCGCGCGC 676  
 Db 1340 TCGGTGGGTGAGGTATCGGCGCTGACGCCAGGCACTGGATTGGAGACGTCCACGA 1399  
 QY 677 CCGGCTGACCGTCTCGAGACGCTCCAGGAGCTGCCAGCAGCGGAGAGCTCGTCTCC 736  
 Db 1400 TCAGG-----GTGCGCGACGCGCTGGAGTACTCGTCCGCGCGGTATGTTGG 1447  
 QY 737 AGTCGCGAGACGCGGAGCGCGCGCGCGCGGTTCAGTTCACCAAGTTCGCTTAC 796  
 Db 1448 GCGCGCGAGACGAAACGCTCCGCGCGCACTGATGCGCGCGCGCGGTGTTGGAGC 1507  
 QY 797 TGCTTACGCCACTCATCGCGGAAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 856  
 Db 1508 GGCTG-----GACTTGTGCAACGAGTGTGTTTTCGTCATACCGATG 1549  
 QY 857 GCGGATGTAAGACGCGCAATTCGCGCGGATCTGGTCAAGGCGTGC-----AGG 910  
 Db 1550 GTGACCGGTGAGTATCCGGGTTTCTGCTGAGGTGTTGATCGGCTGTTGAGAGG 1609  
 QY 911 AAGCGCGGCTTCGGGCTTACGATTCAGATTCGCGGACATCTACGCGGATCTGCTGA 970  
 Db 1610 CTGCTCTGGTTCGCGCGCTTACTCCGACGATCTCGCGGACATCTGCGGCGCGTGG 1669  
 QY 971 TTTCTGCGCGGCTCGCTGCGCGATCTCCGCGCGCTCGCTGCTGCTGCTGCTGCTG 1030  
 Db 1670 TAACGGCGCGGACACCGGTGACGATGTTGTTCTGCCAGCTGGGTGATGATGATTC 1729  
 QY 1031 TCAGGATCTGCTGACGGGCACTG 1056  
 Db 1730 TCAGGTTGACACGTACACGATGTG 1755

RESULT 9  
 AF030986 2404 bp DNA linear PHG 11-JAN-1999  
 LOCUS  
 DEFINITION Mycobacterium phage M56 integrase (int) gene, complete cds.  
 ACCESSION AF030986  
 VERSION AF030986.1 GI:2654006  
 KEYWORDS  
 SOURCE Mycobacterium phage M56  
 ORGANISM Mycobacterium phage M56  
 Viruses.  
 REFERENCE 1 (bases 1 to 2404)  
 Freitas-Vieira, A., Anes, E. and Moniz-Pereira, J.  
 The site-specific recombination locus of mycobacteriophage M56  
 determines DNA integration at the tRNA (Ala) gene of Mycobacterium  
 spp  
 Microbiology 144 (Pt 12), 3397-3406 (1998)  
 JOURNAL  
 MEDLINE 99098701  
 PUBMED 9894232  
 REFERENCE 2 (bases 1 to 2404)  
 Vieira, A., Anes, E. and Moniz-Pereira, J.  
 Mycobacterium phage M56 attP-int sequence  
 TITLE Unpublished  
 JOURNAL  
 REFERENCE 3 (bases 1 to 2404)  
 Vieira, A., Anes, E. and Moniz-Pereira, J.  
 Direct Submission  
 TITLE  
 JOURNAL Submitted (23-OCT-1997) Microbiology, Faculty of Pharmacy

University of Lisbon, Av. Forcas Armadas, Lisbon 1600, Portugal  
 Location/Qualifiers

FEATURES  
 source 1. 2404  
 /organism="Mycobacterium phage M56"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:65388"

repeat\_region

480..502  
 /rpt\_type=inverted

misc\_feature

599..624  
 /note="similar to mycobacterial common core region of  
 homology between attP and attB; similar to 3' end of  
 tRNA-Ala of mycobacteria; these similarities allow the  
 mycobacterium tRNA-Ala gene to remain intact after  
 prophage integration"

gene

697..1817  
 /gene="int"

CDS

697..1815  
 /gene="int"

/note="recombinase"

/codon\_start=1

/transl\_table=11

/product="integrase"

/protein\_id="AAD03774.1"

/db\_xref="GI:2654007"

/translation="MASVRRVRKDGTTAYLVSYRFGGSGAAGATFTDNKRAADAF  
 AAVDAHGAALAHMGHINPAPRTKSELTAIVAEIRHIDHITGVEQYTDKYEYLAN  
 DIBNLGDIPLSLSEEDIAWVWVETTCGRDNGHAPKTLRNKYGLSGALNAVP  
 RYLSNPASGERLPRNGAEDDIRMLTAEFDRLDAPTHWKLAVQFMSVGLRWG  
 EVSALQPRHVDLETSTIRVQAKWYSAGVLPKTKSRRTVDVPALELLRLDLSN  
 EFWNTDGGPVRYPGRVWNPFAVEKAGLVERPFPDHLRHTYASWQLTGSTPTIV  
 SRQLGHESIQTIVTDYTDRTSSRVAAEFMDGLLGF"

1825..1861

/rpt\_type=inverted

ORIGIN

Query Match 5.2%; Score 61.6; DB 7; Length 2404;

Best Local Similarity 48.5%; Pred. No. 12; Indels 51; Gaps 6;

Matches 391; Conservative 0; Mismatches 364;

257 TCCGCAACCACTCTCTCCATCTCGGCCATCTCACCTTGACAGCTGAGCGGCGAG 316

995 TCGCAACGACATCAACCAACCTCGCGACATCCCTTGTGGAAGCTCTCGAAGAG 1054

317 TCACCCAGAGTGGTCAACACCTGAGAGCGCGCGTGGCGCGTGGCGAGTCCACGC 376

1055 ACATCGCCCGTGGGTGAAGTCAAGAAACACCGCGCGCG-----CGACGGCAACGG 1109

377 GGGGTCTCGGAAGCGCTGGCAGCAAGACATCAGCACTGCCACGGCTGTGCAACA 436

1110 GCACGCCCGAAACCCCTCCGCAACAAATACGG---GTTCCTATCGGGGCACTGAACGC 1166

437 CGATCTCGCGCGCGGATCGCGCGGAAAACGATCAGGCTCAACCGTCTCTTCGACGA 496

1167 CCGCGTCCCGCGATCTTGTCCACCAACCTCGCTCGGGCGCGCGCTGCGCGCGGAA 1226

497 TGCTGCCCCGCGGAGCGGAAAGAGATGAAGTTTCCTGAGCGACCCGAGATCGGTCCGC 556

1227 CGCTG-----AGGACGACGACGATCCGATGCTCACCGCGGAGTTCGACCGCGC 1279

557 TTATCAGCGCTTCGCGCGACCTGGGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 616

1280 TCCGCGACCGGTGACACCTCACTGGAAGCTGATGGTTCACTTTCATGGTGTGACCGGTT 1339

617 TGAGTGGGTGAGCGGATCGGCTGCGCGCGCGCGCGGCTGCTGCTGCTGCTGCTGCTG 676

1340 TCGGTGGGTGAGGATATCGCGCTGCGAGCCAGCATGTGATTTGGAGACGTCCACGA 1399

677 CCGGCTGACCGTCTGCTGAGCAGGTCCAGGAGCTGGCGAGAGTGGGAGAGTCTCTTCC 736

1400 TCAGG-----GTGCGGAGGCGGTGGAAGTACTCGTCCGCGCGGTGATGTTGG 1447

737 AGTCCCGAAGACCGGAGGCGCGCGCGCGCGGTTCAGTTCACCGAAGATCGCTCTAC 796

```
Db 1448 GCGCCGCGGAGAGAAACGGTCCCGCCGACCGTGGATGTGCGGCCAGGTGTTGTTGGAGC 1507
QY 797 TGTATTACGCACTCATCTCGCGGAAAGAAAGTACAGAGTGTGTTTACCGCGCGGAAAG 856
Db 1508 GGCTG-----GACTTGTGCAACGAGTTTGTTCGTCAATACCGATG 1549
QY 857 GCGGATGTAGAGCGCGCAATTTCCGCGGATCTGGTCAAGCGGTGCG-----AGG 910
Db 1550 GTGGACCGGTGAGGTATCCGGGGTTTCTCGTAGGTTGTGGAATCCGGCTGTGGAGAAG 1609
QY 911 AAGCGGGCTTCGGGCTTACGCATTACGATCTCGGCACACTCACGGCGGATCCTGA 970
Db 1610 CTGGTCTGTTCCGCGGCTACTCCGACGATCTCGGCACAGTACCGGTCTGGGAGC 1669
QY 971 TTTTTCGCGGGCTCGCTGTGCGGGAATCTCCGCGGCTCGGTCACTCGTCAATCGCGG 1030
Db 1670 TAACGGCGCGGACACCGGTGAGGATTTGTCTCGCAGCTGGTCAATGAGTCAATCAGA 1729
QY 1031 TCACGGATCTGTGTACGGGACCTG 1056
Db 1730 TCACGGTGACAGTACACGGATGTG 1755

RESULT 10
AX655393/c
LOCUS 2000 bp DNA linear PAT 22-MAR-2003
DEFINITION Sequence 5263 from Patent WO03000898.
ACCESSION AX655393
VERSION AX655393.1 GI:29158207
KEYWORDS Oryza sativa
SOURCE Oryza sativa
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE
1 Chang, H.S., Chen, W., Cooper, B., Glazebrook, J., Goff, S.A., Hou, Y.M.,
Katagiri, F., Quan, S., Tao, Y., Whitham, S., Xie, Z., Zhu, T. and Zou, G.
Plant genes involved in defense against pathogens
Patent: WO 03000898-A 5263 03-JAN-2003;
Syngenta Participations AG (CH)
Location/Qualifiers
1. 2000
/organism="Oryza sativa"
/mol_type="unassigned DNA"
/db_xref="taxon:4530"

FEATURES
source
ORIGIN
Query Match 5.1%; Score 60.4; DB 6; Length 2000;
Best Local Similarity 10.7%; Pred. No. 19;
Matches 65; Conservative 277; Mismatches 263; Indels 3; Gaps 1;

QY 204 GAGCTACGAAAGACGCTGAAACCCGACCGCTGAACTCGGAGGCAACCGATCCGCA 263
Db 653 GARSWMYKYSKSAKCKKTRTWTSTSYMTGMYSSYKSNWTSKYSYKGTCTWYT 594
QY 264 CCACCTTCCTGCCATCTCTCCCTTACAGAGTGTGGACGGGAGGTCAACCA 323
Db 593 SNKGSSTRSKMGWSGMSRMWNRKWKRYMYRMKXCTWRRCMYRWGYTMYTTSR 534
QY 324 GCAGTGGGTCAAC---GACCTGGAGCGCGCTCGCGCGGCTGGCGGAGTCCAGCGGG 380
Db 533 SRMYTGRYKARYTKRRYMYTKRYMYTYGMYMKSYMYRYGYCKAKKKCYANCW 474
QY 381 TCGTCGAGCGCGCTGGCAGCAAGACGATCAGCAACTGCCACGCGCTGCTGCACACAT 440
Db 473 KAAYSMMWYRYKYSKWRMTSKYMSWYKCRSKYKAGCYGCKRWYVCSYGYMKW 414
QY 441 CTGCGCGCGGCGATCGCGGCAACGATCAGGCTCAACCGTCTCTTCGACGATGCT 500
Db 413 YTMGSYKYSRCYKRYMYKGMWYMYYSAYSSMTWYIYAKYKRYKRGRTMSWY 354
QY 501 GCCCGCGCGGCGGCAAGAGATGAAGTCTCTGAGACCGCGGAGATCGGTCCGCTTAT 560
```

```
Db 353 GKSYKKYCTWYKMYKMRCYRWRKMRKKTKYSKRCYCWRYATCYWCCCYRKGWYSR 294
QY 561 CACGCGCTTCGCGCGCACTGGGACCGCTCGTCATGCTGTGTGTGGCGACCGGTCTGAG 620
Db 293 SMRTAGKMRWSRWCRSYSWYKTKWKKSYWYSYKWSRSGTWSSAAKRYKY 234
QY 621 GTGGGTGAGCGGATCGGCTCGCGCGCGCGCGGTGCGACCTGCTCGCGCGCGCGCG 680
Db 233 STSRRAKMRACRMYSAACRRYSRTSYCGSGSYGSSKWKYMSKSCSRMTCSWCSCT 174
QY 681 GCTACCGCTGTGCGAGCAGCTCCAGGAGCTGGCAGCAGCGGAGAGCTGCTTCCAGTC 740
Db 173 CYYGAMCWSGMSYMGSCGCTGTGWRKSKYCKKCYCCTCTKYSYTYIRYCKWY 114
QY 741 GCCAAGACCGGAGCGCGCGCGCGCGGTGCGAGTTCACACGAAAGTCTCTACTGCT 800
Db 113 KYSYKCYCYCYWYMYMYMKCMCSRSCSSWMSCAYCTSTSTSRWMSYAAKMGMC 54
QY 801 TACGCCAC 808
Db 53 GSSGMYRM 46

RESULT 11
AC130981
LOCUS 231001 bp DNA linear HTG 13-MAY-2003
DEFINITION Rattus norvegicus clone CH230-229C14, WORKING DRAFT SEQUENCE.
ACCESSION AC130981
VERSION AC130981.3 GI:30578466
KEYWORDS HTG; HTGS PHASB2; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 231001)
Muzny, D., Marle, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Angulano, D.,
Ayalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Chen, Z., Chen, J.,
Chacko, J., Chavez, D., Chen, R., Chen, R., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Henderson, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshewa, L., Lousheed, H., Lozano, R., Lu, X., Lu, X., Ma, J.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwachukwu, O., Okwunodu, G., Olarnpunsoo, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
```



Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quirroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P., Rives, C., Rodley, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.O., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, J., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

# TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL

## REFERENCE AUTHORS TITLE JOURNAL

### COMMENT

On Oct 10, 2002 this sequence version replaced gi:21908142. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Center: Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GXZM  
Center clone name: CH230-94P3  
----- Summary Statistics

Assembly program: Phrap; version 0.990329  
Consensus quality: 234624 bases at least Q40  
Consensus quality: 237798 bases at least Q30  
Consensus quality: 239372 bases at least Q20  
Estimated insert size: 241391; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 258319; contig of 258319 bp in length.  
\* Location/Qualifiers  
1..258319

#### FEATURES source

/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-94P3"  
1..1335  
misc\_feature  
clone end:T7"  
/note="wgs\_end\_extension  
3017..68966  
/note="clone\_boundary  
clone end:T7"  
site:ECORI  
end sequence:BH356710"  
3050..3959  
/note="clone\_boundary  
clone end:T7"  
site:ECORI  
end sequence:BH356710"  
12645..13876  
/note="wgs\_end\_extension  
clone end:T7"  
250432..251175  
/note="clone\_boundary  
clone end:Sp6  
site:ECORI  
end sequence:BH356743"  
256123..258319  
/note="wgs\_end\_extension  
clone end:T7"  
ORIGIN  
Query Match 5.18; Score 59.8; DB 2; Length 258319;  
Best Local Similarity 63.68; Pred.No. 7.4; Indels 0; Gaps 0;  
Matches 91; Conservative 0; Mismatches 52;  
QY 1029 GGTGACGAGTCTGCTGTACGGGCGACCTGCGTGAGGAGTCTGCGGGGTCGCGGC 1088  
DB 121339 GGTGAGGAGGAGATGAGGAGGAGATGAGGAGGAGATGAGGAGGAGATGAGGAGGAGGA 121398  
QY 1089 GATGACGAGGAGTGGCGGCTCGGCTGAGGACCTGAGGCGGAACCTGACGAGGA 1148  
DB 121399 GATGAGGAGGCGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA 121458  
QY 1149 GCTGACGAGCTGTGTTGCCGACG 1171  
DB 121459 GATGAGGAGTGTGATGAGGAGG 121481  
RESULT 13  
PM7G11B/c  
LOCUS PM7G11B 1094 bp DNA linear STS 29-MAY-2003  
DEFINITION Penicillium marneffei STS, clone pm7g11.b, sequence tagged site.  
ACCESSION AL685196  
VERSION AL685196.1 GI:19337723  
KEYWORDS STS.  
SOURCE Penicillium marneffei  
ORGANISM Penicillium marneffei  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.  
1  
Yuen, K.Y., Pascual, G., Wong, S.S., Glaser, P., Woo, P.C., Kunst, F.,  
Cal, J.J., Cheung, E.Y., Medigue, C. and Danchin, A.  
Exploring the Penicillium marneffei genome  
Arch. Microbiol. 179 (5), 339-353 (2003)  
12640520  
PUBMED  
2 (bases 1 to 1094)  
Danchin, A. and Pascual, G.  
Direct Submission  
AUTHORS  
TITLE

JOURNAL Submitted (08-MAR-2002) Danchin A., HKU-Pasteur Research Centre,  
Dexter HC Man Building 8, Sassoon Road, Pokfulam, Hong Kong

REFERENCE Kropinski, A.M., Huang, V.Y.-F., Kim, D. and Yung, E.

AUTHORS Direct Submission

TITLE Submitted (26-APR-2000) Microbiology and Immunology, Queen's  
University, Room 741, Bottell Hall, Kingston, Ontario K7L 3N6,  
Canada

JOURNAL Unpublished

2 (bases 1 to 3689)

2 Kropinski, A.M., Huang, V.Y.-F., Kim, D. and Yung, E.

Direct Submission

Submitted (26-APR-2000) Microbiology and Immunology, Queen's  
University, Room 741, Bottell Hall, Kingston, Ontario K7L 3N6,  
Canada

FEATURES

source

1. .3689

/organism="Integration vector pVH-2"

/mol\_type="other DNA"

/db\_xref="taxon:135519"

/focus

1. .2402

/organism="Cloning vector pGEM3zf"

/mol\_type="other DNA"

/db\_xref="taxon:135522"

2402. 3689

/organism="Pseudomonas phage D3"

/mol\_type="other DNA"

/db\_xref="taxon:31535"

389. .1249

/gene="bla"

389. .1249

/gene="bla"

/codon\_start=1

/evidence=experimental

/transl\_table=11

/product="beta-lactamase"

/protein\_id="AAG01407.1"

/db\_xref="GI:9885347"

/translation="MSIQHPRVALIPFFAFCPLPVFAHPETLVKVKDAEDQLCARVGY  
IEIDSKKILESFRRPEEPFMWSTFKVLGGLVLSIDAGQOLGRRIHYSDNDLVE  
YSPVTKHLTDGMTRELCSAINTSNTAANLLITIGPKELTFLHNGDHYTEL  
DRPELNEAIPNDRDITPMVAMPTTLRLKLTGELUULASRQQLIDWMEADKVAQPL  
LRGALPAGWFIADKSGAGERSGRIIAALPGDKPSRIYVITYTGSQATMDERNRQIA  
EIGASLTKHW"

2404. .3513

/gene="int"

2404. .3513

/gene="int"

/evidence=experimental

/transl\_table=11

/product="integrase"

/protein\_id="AAG01406.1"

/db\_xref="GI:9885348"

/translation="MPYKNDISAYWISFKSATGKLVRRSSGTADYSAKALEQBRRA  
KAWKEKMGVNPPTFEVITPILQAHQHSYETTVHRIKELREFFAGRVVNDLGG  
QDIYGAHRLDAGASPATINRELAALSAIHNCTELEWALPNPKVRKRRREGRD  
RWLIRARVEALCRAARGQKFGPMLEDFIRLVNTGCRREMLGLEWRRVDPANRLIYL  
EASHTKAGKRSIPINEGMAALKRMAFRSETSPECWVFARANGDRVISLSAGFKQ  
ACQAAKTADPTIHLSTCAALVLSAGVPLADVRDLHGSHSTVAMTERYAHLA PARVD  
AVGVLDVVRREGRSRSVHADNPALHGGPLKLVNT"

ORIGIN

Query Match 5.0%; Score 59.4; DB 12; Length 3689;  
Best Local Similarity 63.8%; Pred. No. 23;  
Matches 90; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 899 AGGCGTGGGAGGACCGGGCTTCGGGCTTACGATTCAGCTCGGCGACACTCAQ 958  
DB 3230 AGGCTGCCAGGACGAGGATTCGGGCTTACGATTCAGCTCGGCGACACTCAQ 958  
QY 959 CGGCGATCCTGATTTCTGCGCGGCGTCCGCTGTGCGGATCTCCGCGGCTCGGTCAC 1018  
DB 3290 CGGCGATCCTGATTTCTGCGCGGCGTCCGCTGTGCGGATCTCCGCGGCTCGGTCAC 1018  
QY 1019 CGTGCATCGGCTGACCGATC 1039  
DB 3350 CGACAGTCGGGATGACTGAAC 3370

JOURNAL Submitted (08-MAR-2002) Danchin A., HKU-Pasteur Research Centre,  
Dexter HC Man Building 8, Sassoon Road, Pokfulam, Hong Kong

REFERENCE Kropinski, A.M., Huang, V.Y.-F., Kim, D. and Yung, E.

AUTHORS Direct Submission

TITLE Submitted (26-APR-2000) Microbiology and Immunology, Queen's  
University, Room 741, Bottell Hall, Kingston, Ontario K7L 3N6,  
Canada

JOURNAL Unpublished

2 (bases 1 to 3689)

2 Kropinski, A.M., Huang, V.Y.-F., Kim, D. and Yung, E.

Direct Submission

Submitted (26-APR-2000) Microbiology and Immunology, Queen's  
University, Room 741, Bottell Hall, Kingston, Ontario K7L 3N6,  
Canada

FEATURES

source

1. .3689

/organism="Integration vector pVH-2"

/mol\_type="other DNA"

/db\_xref="taxon:135519"

/focus

1. .2402

/organism="Cloning vector pGEM3zf"

/mol\_type="other DNA"

/db\_xref="taxon:135522"

2402. 3689

/organism="Pseudomonas phage D3"

/mol\_type="other DNA"

/db\_xref="taxon:31535"

389. .1249

/gene="bla"

389. .1249

/gene="bla"

/codon\_start=1

/evidence=experimental

/transl\_table=11

/product="beta-lactamase"

/protein\_id="AAG01407.1"

/db\_xref="GI:9885347"

/translation="MSIQHPRVALIPFFAFCPLPVFAHPETLVKVKDAEDQLCARVGY  
IEIDSKKILESFRRPEEPFMWSTFKVLGGLVLSIDAGQOLGRRIHYSDNDLVE  
YSPVTKHLTDGMTRELCSAINTSNTAANLLITIGPKELTFLHNGDHYTEL  
DRPELNEAIPNDRDITPMVAMPTTLRLKLTGELUULASRQQLIDWMEADKVAQPL  
LRGALPAGWFIADKSGAGERSGRIIAALPGDKPSRIYVITYTGSQATMDERNRQIA  
EIGASLTKHW"

2404. .3513

/gene="int"

2404. .3513

/gene="int"

/evidence=experimental

/transl\_table=11

/product="integrase"

/protein\_id="AAG01406.1"

/db\_xref="GI:9885348"

/translation="MPYKNDISAYWISFKSATGKLVRRSSGTADYSAKALEQBRRA  
KAWKEKMGVNPPTFEVITPILQAHQHSYETTVHRIKELREFFAGRVVNDLGG  
QDIYGAHRLDAGASPATINRELAALSAIHNCTELEWALPNPKVRKRRREGRD  
RWLIRARVEALCRAARGQKFGPMLEDFIRLVNTGCRREMLGLEWRRVDPANRLIYL  
EASHTKAGKRSIPINEGMAALKRMAFRSETSPECWVFARANGDRVISLSAGFKQ  
ACQAAKTADPTIHLSTCAALVLSAGVPLADVRDLHGSHSTVAMTERYAHLA PARVD  
AVGVLDVVRREGRSRSVHADNPALHGGPLKLVNT"

ORIGIN

Query Match 5.0%; Score 59.4; DB 11; Length 1094;  
Best Local Similarity 43.9%; Pred. No. 31;  
Matches 289; Conservative 0; Mismatches 367; Indels 2; Gaps 2;

QY 106 GCGATGGTGCAGTTCGTCGGGAGTGTGTCGCGAGTACGAAAGACGTGAAA 225  
DB 838 GCGNCCGCGCGCGTTCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 779  
QY 166 CAGATTACCTCGCCGATTTCTGCGGGAGTGTGTCGCGAGTACGAAAGACGTGAAA 225  
DB 778 CCG 719  
QY 226 CCGACCGCGTGAACTCGAGGAGCAACCGATCCGCAACACCTCTCTGCCATATCTCGC 285  
DB 718 CCG 660  
QY 286 CATCTCACCTTACGAGCTGACGCGGAGTGCACCCAGCAGTGGGTCAACGACCTGGAG 345  
DB 659 CCG 600  
QY 346 GCGCGCGTGGCGCGTGGCGGAGTCCACGCGGGGTCTCGGAGCCGCTGCACGCGAG 405  
DB 599 GCG 540  
QY 406 ACATCAGCAACTGCCACGCGCTGTGTGACACAGATCTGCGCGCGCGCGCGCGAA 465  
DB 539 GCG 480  
QY 466 CCGATCAGGCTCAACCGTGTCTTTCGACGATGTCTCCCGCGCGCGCGCGCGAGATG 525  
DB 479 GCG 420  
QY 526 AAGTTCTGACGACCGCGGAGTCTGCGCGCTTATCACGCGCTTCCGCGCGACTGGCGA 585  
DB 419 GCG 360  
QY 586 CCGCTCTGTCATGCTGTGTGGCGACCGCGTCTGAGGTGGGGTGAAGCGATCGGCTGCGC 645  
DB 359 CCG 300  
QY 646 GCGCGCGCGTGAACCTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 704  
DB 299 CCG 240  
QY 705 GCGAGTGGCGGACGCGGAGAGTCTCTTTCAGTGGCGCGCGCGCGCGCGCGCG 762  
DB 239 GCG 182

RESULT 14

AF260831 3689 bp DNA circular SYN 23-AUG-2000

LOCUS Integration vector pVH-2 complete sequence.

DEFINITION Integration vector pVH-2

ACCESSION AF260831

VERSION AF260831.1 GI:9885345

KEYWORDS Integration vector pVH-2  
Integration vector pVH-2  
artificial sequences; vectors.

ORGANISM Kropinski, A.M., Huang, V.Y.-F., Kim, D. and Yung, E.  
Characterization of the int-att region of bacteriophage D3, and the  
attD3 site on the Pseudomonas aeruginosa genome

RESULT 15  
 AF165214/C  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Pseudomonas phage D3  
 Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;  
 Lambda-like viruses.  
 1 (bases 1 to 56425)  
 Farinha,M.A., Allan,B.J., Gertman,E.M., Ronald,S.L. and  
 Kropinski,A.M.  
 Cloning of the early promoters of Pseudomonas aeruginosa  
 bacteriophage D3: sequence of the immunity region of D3  
 J. Bacteriol. 176 (16), 4809-4815 (1994)  
 94327444  
 8050993  
 2 (bases 1 to 1220)  
 Sharp,R., Jansons,I.S., Gertman,E. and Kropinski,A.M.  
 Genetic and sequence analysis of the cos region of the temperate  
 Pseudomonas aeruginosa bacteriophage, D3  
 Gene 177 (1-2), 47-53 (1996)  
 97080499  
 3 (bases 53297 to 54321)  
 Kropinski,A.M. and Sibbald,M.J.  
 Transfer RNA genes and their significance to codon usage in the  
 Pseudomonas aeruginosa lamboid bacteriophage D3  
 Can. J. Microbiol. 45 (9), 791-796 (1999)  
 99455323  
 10526403  
 4 (bases 1 to 56425)  
 Gilakjan,Z.A. and Kropinski,A.M.  
 Cloning and analysis of the capsid morphogenesis genes of  
 Pseudomonas aeruginosa bacteriophage D3: another example of protein  
 chain mail?  
 J. Bacteriol. 181 (23), 7221-7227 (1999)  
 20042341  
 10572124  
 5 (bases 1 to 56425)  
 Kropinski,A.M.  
 Sequence of the genome of the temperate, serotype-converting,  
 Pseudomonas aeruginosa bacteriophage D3  
 J. Bacteriol. 182 (21), 6066-6074 (2000)  
 20485557  
 11029426  
 6 (bases 26775 to 28246)  
 Kropinski,A.M., Huang,V.Y.-F., Kim,D. and Yung,E.  
 Characterization of the int-att region of Bacteriophage D3, and the  
 attD3 Site on the Pseudomonas aeruginosa genome  
 Unpublished  
 7 (bases 43335 to 46477)  
 Farinha,M.A., Allan,B.J., Gertman,E.M., Ronald,S.L. and  
 Kropinski,A.M.  
 Direct Submission  
 Submitted (04-AUG-1993) Microbiology, Queen's University, Room  
 741/743, Botterell Hall, Kingston, Ontario K7L 3N6, Canada  
 8 (bases 46478 to 46764)  
 Kropinski,A.M.  
 Direct Submission  
 Submitted (31-JUL-1995) Microbiology, Queen's University, Room  
 741/743, Botterell Hall, Kingston, Ontario K7L 3N6, Canada  
 9 (bases 1 to 1220)  
 Kropinski,A.M. and Sharp,R.W.  
 Direct Submission  
 Submitted (29-JAN-1996) Microbiology, Queen's University, Room  
 741/743, Botterell Hall, Kingston, Ontario K7L 3N6, Canada  
 10 (bases 53297 to 54321)  
 Kropinski,A.M. and Sibbald,M.J.  
 Direct Submission  
 Submitted (08-JUL-1998) Microbiology, Queen's University, Room

741/743, Botterell Hall, Kingston, Ontario K7L 3N6, Canada  
 11 (bases 1 to 7657)  
 Kropinski,A.M. and Gilakjan,M.A.  
 Direct Submission  
 Submitted (04-MAY-1999) Microbiology, Queen's University, Room 743,  
 Botterell Hall, Kingston, Ontario K7L 3N6, Canada  
 12 (bases 26775 to 28246)  
 Kropinski,A.M., Huang,V.Y.-F., Kim,D. and Yung,E.  
 Direct Submission  
 Submitted (03-JUL-1999) Microbiology, Queen's University, Room 743,  
 Botterell Hall, Kingston, Ontario K7L 3N6, Canada  
 13 (bases 1 to 56425)  
 Kropinski,A.M.  
 Direct Submission  
 Submitted (05-JUL-1999) Microbiology & Immunology, Queen's  
 University, Room 741, Botterell Hall, Kingston, Ontario K7L 3N6,  
 Canada  
 On or before Jul 2, 2000 this sequence version replaced gi:403435,  
 gi:984852, gi:1685011, gi:4927380, gi:5059247, gi:6166381.  
 FEATURES  
 Location/Qualifiers  
 1..56425  
 /organism="Pseudomonas phage D3"  
 /mol\_type="genomic DNA"  
 /specific\_host="Pseudomonas aeruginosa"  
 /db\_xref="taxon:31535"  
 37..417  
 /genes="orf1"  
 37..417  
 /genes="orf1"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="Orf1"  
 /protein\_id="AAD38953.1"  
 /db\_xref="GI:5059248"  
 /translation="MARHKQPAEVAKFGADKKNPORYOEPAKGRGDIGEAPIHLQG  
 PARLAKKEICTQSIKGVILGSLTLEVTANLLAEYADPTFEAVGKYTHLIGNLRL  
 GLTPSDRQKFGLEKPKXKDEFFDF"  
 419..2110  
 /genes="orf2"  
 419..2110  
 /genes="orf2"  
 /note="similar to Escherichia coli ymfN"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="terminase"  
 /protein\_id="AAD38954.1"  
 /db\_xref="GI:5059249"  
 /translation="MTPSDIASQYASDVVGAIACVYVALACORELNDLRQDDDDW  
 PYYDAKADRAVKFQMLPHYKQKWSASKLVFEPQVFIENIFGVKRTGKR  
 FREAYBEIPKNGKSARLAARGIYLPADGESGAETVSGATTEKQAFVFRPAMWNAH  
 KLENLRRFGIELSGNQKPMFVMDMSKFTVIGNFGDGSAAALVDVEYEHDT  
 DALVDTMOTGMGAREQFLSIITTAGSNLGGPCHEKRRDVRILEQGTIDETLFGIY  
 TIDEDDPDDPASLIKANPNYGVSPFDLLAQLOAKRSASQKNAFRFKQWVGA  
 RTVMNMLAWOROKRDTTIADMGACRCWMLDLASKVAALVLMFEKAGQFYCIPIRF  
 YAPAAAEENKYQNEALEGHLVLTGSMTDYAFIADILDLAKQIDLQDAFDDQQA  
 NYILRLSNTSIPVDFNQTIVKMSDPMKEVEARVARTLWHGNGPMTWMMGNVAAK  
 IDAKENIYRKENDPNCKIDGFTVLIWAGRALVAGVDDGDDFMNAINPIIA"  
 2107..2271  
 /genes="orf3"  
 2107..2271  
 /genes="orf3"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="Orf3"  
 /protein\_id="AAF80770.1"  
 /db\_xref="GI:8895112"  
 /translation="MNIACTGLYFFGVGLGALFVACTFVLLGLGWLISGASAPAFIA  
 AFIRKGLTSE"  
 2284..3358  
 /genes="orf4"

Search completed: August 6, 2004, 11:53:05  
Job time : 5421.76 secs

```
CDS
2264. .3568
/gene="orf4"
/notes="similar to Escherichia coli bacteriophage HK97
portal protein: SwissProt Accession Number P49859"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="portal protein"
/protein_id="AAD38955.1"
/db_xref="GI:5059250"
/translation="MSKSGKVLVSATSAPRSLFGWGKTIIRLTGDAFWQSLGREGS
SGCKVTVDKAVKSAVWACVRLISTVAGLEPLGYERKADGSRVDARSFLYDVVHN
SPNDDMTAFQFQWAMVAGMLLWGNAYASIRRAAGPAAALDFLLPSRVLECDENGRLK
YFTTKGARRIERITNMLHI PAFTLDGRIGLSAIRYGVDFVGSVMSAEADAANGTFKN
GLLPTVAFKVRILLOPAOREPEREVKVSNGAMSGRSPVLEQGITPETIGINPVDAQ
LLETRHGVIETCRNFGVPFMWIGTGDGSGNWGTGLEQMLAFLFESSITNOIQOC
VNRLLUTAPERIRYAEFSLGLEFLKADAGRAAWYSTWAQGFMTNRGRKKNLPEL
PGDILTVQSNLIVPIDQLGQSNKQAVRAALMNWFSQPEPQE"
3572. .4462
/gene="orf5"
3572. .4462
/gene="orf5"
/notes="similar to Streptomyces coelicolor ClpP protease
encoded by GenBank Accession Number AF071885"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="ClpP protease"
/protein_id="AAD38956.1"
/db_xref="GI:5059251"
/translation="MILRLPAAPVRPRSGVQCDLAPKALDAMWPELRAAGDNPD
TITIEPIGYDWTGEGVTAKRIAGALRSIGNDVTNINSPPGDFEGEGLAIYNLLR
EKKGVTNIIIGLAASAFIAGAGDEIRIGSAFLMIHNAWLIANGNRLREIADW
LEPFDMTLADIYQAQTEIDIDIVKMDAETWIGGREAVDKGWADAFLESDEISSAPS
NRSEATLAKRRNDALARSMPRSQRNELINDFKTSMLEAAGGGGDTPTDMPGAVAP
LSAALPAQDITKFLQGESQ"
4459. .5646
/gene="orf6"
4459. .5646
/gene="orf6"
/notes="similar to Escherichia coli bacteriophage HK97
major head protein: SwissProt Accession Number P49861;
proteolytically processed and cross-linked in mature phage
particles"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="major head protein"
/protein_id="AAD38957.1"
/db_xref="GI:5059252"
/translation="MSDFKQIGELNASLKQVGDQIKSQAEQVNTQIANFGEMNKETR
AKVDLLTAGLQARLSAAEQMLANERKDGGEAPKTAGQWVAESKEQGVTSLLR
GSHRVSMPSRAISIDSGCALVAPDRPGVVAAPORRLTIDLVAPTTESNVEYV
RTGFGVNAAPVSEGTQKPYSDLTFELENAPVRTIAHLFKASRQLDDASALQSYIDA
RARYGLMVEECQLLYNGTGANLHGIIPQAQYAPPGVVVTAQRIDRIRLAILOA
QLAEFPASGIVLNPIDWALIENLKDANRYIIGSPQNGTPTLWELPVVETQAITODE"

gene
CDS
Query Match 5.0%; Score 59.4; DB 7; Length 56425;
Best Local Similarity 63.8%; Pred. No. 12;
Matches 90; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
QY 899 AGCGTGGCAGGAGCGGGCTCCCGGGCTTACGATTCACGATCTCGGACACCTCAG 958
|||||
Db 27386 AGSCCTGCCAGGAGCGAGATTGGCGACTTTACGATTCACGACCTCGGCCACCTCGG 27327
|||||
QY 959 CGCGGATCTGATTTCTGCGGGCGTCCGCTGTGCGCGATCTCCCGCCGCTCGGTCACT 1018
|||||
Db 27326 CGGCATGGTGTGTCAGCGCGCGGCTTCGCTGCGGATCTTCGGATCTGCTCGGACACT 27267
|||||
QY 1019 CGTCGATCGCGTCAACGATC 1039
|||||
Db 27266 CGACAGTCGGGATGACTGAAC 27246
|||||
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2004, 09:31:14 ; Search time 141.71 Seconds  
(without alignments)  
4617.079 Million cell updates/sec

Title: US-09-855-340A-1

Perfect score: 1179

Sequence: 1 gtgtggatcgagaagacgg.....tgtggcgacgacgatga 1179

Scoring table: IDENTITY NUC

Gapop 10\_0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:\*

1: /cgn2\_6/ptodata/2/ina/5A COMB.seq:\*

2: /cgn2\_6/ptodata/2/ina/5B COMB.seq:\*

3: /cgn2\_6/ptodata/2/ina/6A COMB.seq:\*

4: /cgn2\_6/ptodata/2/ina/6B COMB.seq:\*

5: /cgn2\_6/ptodata/2/ina/PTUS COMB.seq:\*

6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	56.6	4.8	1161	4	US-09-821-167-15
2	56.6	4.8	11188	4	US-09-821-167-1
3	51.4	4.4	516	4	US-09-183-861-34
4	51.4	4.4	516	4	US-09-022-765-34
5	51.4	4.4	516	4	US-09-551-974A-34
6	51.4	4.4	516	4	US-09-565-501A-34
7	51.4	4.4	516	4	US-09-639-206A-34
8	51.4	4.4	516	4	US-09-874-923-34
9	51.4	4.4	582	4	US-09-874-923-114
10	51.4	4.4	7065	4	US-09-874-923-115
11	50.4	4.3	47981	4	US-09-679-279-1
12	50.2	4.3	534	4	US-09-252-991A-11386
13	50.2	4.3	3150	4	US-09-252-991A-11487
14	50.2	4.3	3372	4	US-09-252-991A-11405
15	48.8	4.1	915	4	US-09-252-991A-2186
16	48.8	4.1	936	4	US-09-252-991A-2345
17	48.8	4.1	1059	4	US-09-252-991A-2490
18	48.6	4.1	744	4	US-09-252-991A-15538
19	48.6	4.1	807	4	US-09-252-991A-15427
20	47.4	4.0	47981	4	US-09-679-279-1
21	46.8	4.0	4403765	3	US-09-103-840A-2
22	46.8	4.0	4411529	3	US-09-103-840A-1
23	46.4	3.9	1896	4	US-09-343-011B-4
24	45.8	3.9	858	4	US-09-252-991A-12882
25	45.8	3.9	897	4	US-09-252-991A-12720
26	45.8	3.9	1398	4	US-09-252-991A-13082
27	45.8	3.9	2145	4	US-09-252-991A-13235

c 28	45.8	3.9	12001	1	US-08-458-568A-11	Sequence 11, Appli
29	45.6	3.9	1176	4	US-09-252-991A-6593	Sequence 6593, Ap
30	45.6	3.9	1440	4	US-09-252-991A-6573	Sequence 6573, Ap
31	45.6	3.9	2322	1	US-09-252-991A-6621	Sequence 6621, Ap
32	45.6	3.9	20235	1	US-07-642-734C-3	Sequence 3, Appli
33	45.6	3.9	20235	1	US-07-642-734C-3	Sequence 3, Appli
34	45.6	3.9	20235	3	US-08-439-009A-3	Sequence 3, Appli
35	45.6	3.9	20235	3	US-08-439-009A-3	Sequence 3, Appli
36	45.2	3.8	813	4	US-09-252-991A-2893	Sequence 2893, Ap
37	45.2	3.8	876	4	US-09-252-991A-2800	Sequence 2800, Ap
38	45	3.8	1296	4	US-09-252-991A-3696	Sequence 3696, Ap
39	45	3.8	1368	4	US-09-252-991A-3813	Sequence 3813, Ap
40	45	3.8	2784	4	US-09-252-991A-3769	Sequence 3769, Ap
41	45	3.8	68750	3	US-09-335-409-1	Sequence 1, Appli
42	45	3.8	68750	4	US-09-568-102-1	Sequence 1, Appli
43	45	3.8	68750	4	US-09-567-969-1	Sequence 1, Appli
44	45	3.8	68750	4	US-09-568-480-1	Sequence 1, Appli
45	45	3.8	68750	4	US-09-568-486-1	Sequence 1, Appli

#### ALIGNMENTS

##### RESULT 1

US-09-821-167-15

; Sequence 15, Application US/09821167

; Patent No. 6569668

; GENERAL INFORMATION:

; APPLICANT: Hosted Jr., Thomas J.

; APPLICANT: Horan, Ann C.

; TITLE OF INVENTION: Isolated Nucleic Acids from Micromonospora rosaria

; TITLE OF INVENTION: Plasmid pMR2 and Vectors Made Therefrom

; FILE REFERENCE: IN011490

; CURRENT APPLICATION NUMBER: US/09/821.167

; CURRENT FILING DATE: 2001-03-29

; PRIOR APPLICATION NUMBER: US 60/194,461

; PRIOR FILING DATE: 2000-04-04

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: Patent In Ver. 2.1

; SEQ ID NO 15

; LENGTH: 1161

; TYPE: DNA

; ORGANISM: Micromonospora rosaria

US-09-821-167-15

Query Match 4.8%; Score 56.6; DB 4; Length 1161;  
Best Local Similarity 46.2%; Pred. No. 0.0034;  
Matches 262; Conservative 0; Mismatches 299; Indels 6; Gaps 2;

Qy	545	AGATCGGTGCGCTTATCAGCGCGCTTCCGCGCGACTGGCGGACCGCTGCTCATGCTGCTGG	604
Db	557	AGTCTGTCGCGCTTGTGAGGCCATACCCGCGTACCGGCGGCGGTCTGTATCGCGG	616
Qy	605	TGGCGACCGGTCTGAGTGGGTGAGCGATCGGCTTGGCGCGCGCGGTCGACCTGC	664
Db	617	CTGTGTCGCGCTTACGCGCGGAGAGATCGAGACTCCGGTCCGAGACGTGACCTG-	675
Qy	665	TCGCGCGCGCGCGCGCTGACCGTGTGAGAGTCCAGAGTGGCGGACCGGAG	724
Db	676	--ACCGAACACACATCACCGTCCGAAAGCCAGGTCGAAACCGCTGCACGACGAGGGA	733
Qy	725	AGTCTGCTTCCAGTCCCGAAGACCGCGCGCGCGCGCGCTCAGTTTACCACGA	784
Db	734	AGCGTTTGACAGGATCCCAAGTCCGAGCGGCGGCAAGCAACATCGCCATCCCTCCC	793
Qy	785	AGTCTGCTTACTGTTTACGCCACTATCCCGGAAAGAAAGTACGAGGTGCTGTCA	844
Db	794	ACGTCTGATACCGGTGATCCGCTCGACCTCGACAGTTCGCGGGAAGGATCGCTGTTCG	853
Qy	845	CCGCGCGGAAAGCGGATGTTAAGACGCGCAATTTCCGCGCGATCTTGGGTCAAGCGGT	904
Db	854	TCAGCGCGAGGATACCCCTCGCGCGGCGACACCTGTACGAGGCTTCTGACGCGCTC	913

QY 905 GCGAGGAAGCGGGCTTCGGGCTTACGATTCAGCATTCGGGCACACTCAGCGGCGA 964  
DB 914 GGGGAAGTCCGATTCGACACCTTCCAGACTTCGGGCACACCGGTACAGCC 973  
QY 965 TCCTGATTTCTCGCGGGCTTCGGCTGTGCGGATCTCCGCGCTCGGTCACTGTCGA 1024  
DB 974 TGCCCGCAGACCGGGGACCTTGCGCGACCTGATGAGCGGTTCGGGCACTGTCGA 1033  
QY 1025 TCGGGTCAAGATTCGTTGACGGGACCTCGGTGAGAGGTCAAGAGGATCTCTCG 1084  
DB 1034 TGCGTG---CGGCTCGCGGTACCTCCAGCGGTTGACGGTCTGACCGGGAGATCGCCA 1090  
QY 1085 CGCGATCGAGGAGCGATGGCGCG 1111  
DB 1091 AAGCCCTTCCAGAGTGGCGGCGCAG 1117  
RESULT 2  
US-09-821-167-1  
; Sequence 1, Application US/09821167  
; Patent No. 6549668  
; GENERAL INFORMATION:  
; APPLICANT: Hosted Jr., Thomas J.  
; APPLICANT: Horan, Ann C.  
; TITLE OF INVENTION: Isolated Nucleic Acids from Micromonospora rosaria  
; TITLE OF INVENTION: Plasmid pMR2 and Vectors Made Therefrom  
; FILE REFERENCE: IN011490  
; CURRENT APPLICATION NUMBER: US/09/821,167  
; CURRENT FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: US 60/194,461  
; PRIOR FILING DATE: 2000-04-04  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 11188  
; TYPE: DNA  
; ORGANISM: Micromonospora rosaria  
; FEATURE:  
; NAME/KEY: RBS  
; LOCATION: (6055)..(6059)  
; NAME/KEY: RBS  
; LOCATION: (6391)..(6394)  
; NAME/KEY: RBS  
; LOCATION: (8084)..(8088)  
; NAME/KEY: RBS  
; LOCATION: (9834)..(9837)  
; NAME/KEY: RBS  
; LOCATION: (10010)..(10012)  
US-09-821-167-1  
Query Match 4.8%; Score 56.6; DB 4; Length 11188;  
Best Local Similarity 46.2%; Pred. No. 0.0043;  
Matches 262; Conservative 0; Mismatches 299; Indels 6; Gaps 2;  
QY 545 AGATCGGTGCGTTATCAGCGGCTTCGCGGCACTCGGCGACCGTCTGTCATGCTGCTGG 604  
DB 10584 AGTCTCGCCCTGCTGAGGACATCAACCCGCGTACCGAGCGGCTCTGATCGCGG 10643  
QY 605 TCGCGACCGGTCTGAGGTGGGTGAGGCGATCGGCTGCGCGCGCGCGGTTCGACCTTCG 664  
DB 10644 CTTGTTGCGGCTTACCGCGGGAGAGATCGCAGGACTCGCGGTGCGACGCTGGACCTG- 10702  
QY 665 TCGCGCGCGGCGCGGCTGACGCTGTCGAGCAGCTCCAGGAGTGGCGACAGCGGAG 724  
DB 10703 --ACCGAACACCATCATCGTCCGGAAGCAGGTCGAACTGACGACGAGGGGA 10760  
QY 725 AGCTCGTCTTCCAGTCCGGAAGACCGGAAGGCGCGCGCGTCAAGTTTCAACCA 784  
DB 10761 AGCGGTTTGAAGAAGTCCCAAGTCCGAGCGGGAAGCAACCATCGCCATCCCTCC 10820  
QY 785 AGTCTGCTTACTGTTAGCCACTCATCGCGGAAGAAAGTACGAGGTGCTGTTCA 844  
DB 10821 ACCTGATCCCGGTGATCCCGCTGACACGAGTTCCCGGGAAGGATCGCTGTTTCG 10890

QY 845 CCGCGCGAAGCGGGATGTTAAGACGCGCAATTTTCGGCGGATCTGGGTCAAGGCGT 904  
DB 10881 TCAGCCGAGCATCACCCCTGCGGCGGACACCCCTGTACAGGGTTCGTACGGCTC 10940  
QY 905 GCGAGGAAGCGGGCTTCGGGCTTACGATTCAGCATTCAGATTCGGGCACACTCAGCGGCGA 964  
DB 10941 GGGGAAGTTCGACTCGACACCCCTCACCTTCACGACCTGCGGCACACCGGTCAAGCC 11000  
QY 965 TCTGATTTCTCGCGGCGTTCGGCTGTGCGGATCTCCCGCGCTCGGTCACTGTCGA 1024  
DB 11001 TCGCGCGCAGACCGGGGACCTTGGCGGACCTGATGAAGCGGCTCGGGCACTGTCGA 11060  
QY 1025 TCGCGGTCAAGGATCTGTTACGGGCACTCGGTGAGGAGTTCGACGAGGAGTCTCTCG 1084  
DB 11061 TGCGTG---CGGCTCGCGGTACTCCAGCGCTTGACGGTTCGTGACCGGAGATCGCCA 11117  
QY 1085 CGCGATCGAGGAGCGATGGCGCGG 1111  
DB 11118 AAGCCCTTTCGAGCTGGCGGCGCAG 11144  
RESULT 3  
US-09-183-861-34  
; Sequence 34, Application US/09183861  
; Patent No. 6365165  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Webb, John R.  
; APPLICANT: Dillon, David C.  
; APPLICANT: Skeiky, Yasir A.W.  
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE THERAPY AND  
; NUMBER OF SEQUENCES: 87  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/183,861  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/022,765  
; FILING DATE: 12-FEB-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.420C3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 34:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 516 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: Leishmania major  
US-09-183-861-34  
Query Match 4.4%; Score 51.4; DB 4; Length 516;  
Best Local Similarity 47.1%; Pred. No. 0.036;

Matches	222; Conservative	0; Mismatches	246; Indels	3; Gaps	2;
QY	302	AGCTGGACGGCAGGTCACCCAGCAGTGGGTCAA	CGACCTGGAGCCGGGGTTCGGCCCGT	361	
Db	7	AGCAGCCCTAGACACCGCCACGACAGCGCCGAGCTGGAGGCACGGGTGGCACCGC	66		
QY	362	GGCCGGAGTCCACGCGGGGTCTGGAGCGCTTGGCAGCGAA	-GAGATCAGCAATGC	420	
Db	67	TGGCCGCGGACCGGACGAGCGCGCCAGCAGCTTGGCCGCGAA	CGCCGAGGAGTGTGAGC	126	
QY	421	CACGGCCTGTGTCACAGCATCTGGCGGCGCGGAGTTCGGCGGAGAAACGATCAGGCTCAAC	480		
Db	127	AGGSCCTAGACACCGCCACGACGACGCGCCGAGCTGGAGGACGGGTGSCACGGCTGG	186		
QY	481	CCGTGCTCTTCGAGGATGCTGCCCGGCGGAGCGGAAAGATGAGTTCTTCTGAGCGAC	540		
Db	187	CCGCGGACGGCAGCAGGCGCCCGCAGCAGTGGCCGCGAA	CGCCGAGGAGCTGACGAGC	246	
QY	541	CCGGAGATCGGTTCGGCTTATCACGGCGGCTTCGCGCGCACTGGCCACCGCTCGTCAATGTG	600		
Db	247	GCCTAGACACCGCCACGACGACGCGCCGAGCTGGAGGCACAGGTGGCACGGCTGGCCG	306		
QY	601	CTGGTGGCGACCGGTCTGAGTGGGGTGAAGCGAT--	CGGCTCGCGCGCGCGCGGTCTCG	658	
Db	307	CGAACCGCGAGGAGCTCAGCAGCGCCCTAGACACCGCCACGACGACGCGCCGAGCTGG	366		
QY	659	ACCTGCTCGCGCGCGCGCCCGGTGACCGCTGCTGAGCAGCTCCAGGAGTGGCCAGCA	718		
Db	367	AGSCACGGTGGACAGGCTGSCCCGCGNACCGCGACAGGCGCGCAGCAGTGGCCCGGA	426		
QY	719	CGGGAGAGTCTGTTCTTCAGTTCGCGAAGACCGCGAAGGGCCGCGCACGG	769		
Db	427	ACGCCGAGGAGCTGCAGCAGCGGCTTAGACACCGGCAACCGCAGCGAGCGGCC	477		

## RESULT 4

US-09-022-765-34  
 ; Sequence 34, Application US/09022765  
 ; Patent No. 6375955  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Reed, Steven G.  
 ; APPLICANT: Campos-Neto, Antonio  
 ; APPLICANT: Webb, John R.  
 ; APPLICANT: Dillon, Davin C.  
 ; APPLICANT: Skeiky, Yasir A.W.  
 ; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USEIN THE THERAPY AND  
 ; NUMBER OF SEQUENCES: 87  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESS: SEED and BERRY LLP  
 ; CITY: Seattle  
 ; STATE: Washington  
 ; COUNTRY: USA  
 ; ZIP: 98104-7092  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IEM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/022,765  
 ; FILING DATE: 12-FEB-1998  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Maki, David J.  
 ; REGISTRATION NUMBER: 31,392  
 ; REFERENCE/DOCKET NUMBER: 210121.420C3  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (206) 622-4900  
 ; TELEFAX: (206) 682-6031  
 ; INFORMATION FOR SEQ ID NO: 34:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 516 base pairs

```

; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Leishmania major
US-09-022-765-34

Query Match      4.4%; Score 51.4; DB 4; Length 516;
Best Local Similarity 47.1%; Pred. No. 0.036;
Matches 222; Conservative 0; Mismatches 246; Indels 3; Gaps 2
Qy 302 AGCTGACGGGACAGGTCAACCCACAGTGGGTCAACACCTGGAGGCGGGCGTGGGCCCGT 361
Db |||
Qy 7 AGCAGCGCTAGACACCGCCACGACGACGCGCCGAGCTGGAGGACACGGTGTGCAGGC 66
Db |||
Qy 362 GGCCGGAGTCCACGCGGGGTCTGTCGAAGCCGCTGGCAGCGAA-GACGATCAGCAACTGC 420
Db |||
Qy 67 TGGCCCGGACCGCAGCAGGCGCGCAGCAGCTGGCCCGAAACGCCAGGAGCTGCAGC 126
Db |||
Qy 421 CAGGGCTGTGTGCACACGATCTCGGCGCGGGCGCATCGCGCGCAAAACGATCAGGCTCAAC 480
Db |||
Qy 127 AGCGCTAGACACCGCCACGACGACGCGCCGAGCTGGAGGACCGGTGGCACGGCTG 186
Db |||
Qy 481 CGGTGCTTTCGACGATGCTGCCCGCGGCGGAGCCGAAAGATGAGTTCTTGAGGGAC 540
Db |||
Qy 187 CCGCGGACGCGCAGCAGGCGCGCCAGCAGCTGGCCCGGAACGCCGAGGAGCTGCAGCAGC 246
Db |||
Qy 541 CCGAGATCGGTGGGCTTATCACGGGCTTCCGCGCACTGGCGACCGCTCGTCTCATGCTG 600
Db |||
Qy 247 GCCTAGACACCGCCACGACGACGCGCGGAGCTGGAGGCACAGGTGGCACGGCTGGCG 306
Db |||
Qy 601 CTGGTGGGACCGGTCTGAGGTGGGTGAGGCGAT--CGGCCCTCGCGCGCGGGGTGCG 658
Db |||
Qy 307 CGAACCGGAGGAGCTGCAGCAGCGCTAGACACCGCCACGACGACGCGCGGAGCTGG 366
Db |||
Qy 659 ACCTGCTCGCGCGCGCGCGCGCTACCGTCTCGTCAGACAGCTCCAGGAGCTGCCCAGCA 718
Db |||
Qy 367 AGGCACGGTGGCACGGCTGGCCCGGACCGCCACGAGCGCGCCAGCTGCCCGGA 426
Db |||
Qy 719 CGGAGAGCTGCTTCTCAGTCCCGGAAGACCCGGAAGGGCCGGCGCACGG 769
Db |||
Qy 427 ACGCCGAGGAGCTGCAGCAGCGCTAGACACCGCCACGACCGCGGAGCTGCCCGCG 477
Db |||

RESULT 5
US-09-551-974A-34
; Sequence 34, Application US/09551974A
; Patent No. 6500437
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Webb, John R.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Sreelky, Yasir A.W.
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
; FILE REFERENCE: 210121.420C5
; CURRENT APPLICATION NUMBER: US/09/551,974A
; CURRENT FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 516
; TYPE: DNA
; ORGANISM: Leishmania major
US-09-551-974A-34

```

Query Match 4.4%; Score 51.4; DB 4; Length 516;  
Best Local Similarity 47.1%; Pred. No. 0.036;  
Matches 22; Conservative 0; Mismatches 246; Indels 3; Gaps 2;  
Qy 302 AGTGGACGGCGAGTCACCACAGTGGGTCAACACTGGAGGCGGCGTGGCCCGT 361

Db 7 AGCAGCGCTAGACACCGCCAGCAGCGCGCGGCTGGAGGACCGGTGGCAGCGC 66  
Qy 362 GCGCGAGTCCAGCGGGTCTGCGAAGCCCTGCGAGCGAA-GACGATCAGCAATGCG 420  
Db 67 TGCCCGGACCGCGAGGCGCGCCAGCAGCTGGCCCGGAAACCGCGAGGAGCTGCAGC 126  
Qy 421 CACGGCTCTGTCACACGATCTGCGCGCGCGGCGATCGCGCGAAGATGAGTCTCAAC 480  
Db 127 AGCGCTTAGACACCGCCAGCAGCAGCGCGCGGCTGGAGGACACGGTGGCAGCGCTG 186  
Qy 481 CGGTGCTCTTACAGATGCTGCGCGCGCGCGCGGCGGCGGAGGATGAAGTCTGAGCGAC 540  
Db 187 CCGCGACCGCGACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 246  
Qy 541 CCGGAGATCGGTGGCTTATCAGCGCGCTTCCCGCGCGCACTGGCGACCGGTCTCATGCTG 600  
Db 247 GCTAGACACCGCCAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 306  
Qy 601 CTGGTGGCGACCGGTCTGAGGTGGGTGAGGCGAT--CGGCTGCGCGCGCGCGCGCGCG 558  
Db 307 CGAAGCGCGAGGAGCTGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 366  
Qy 659 ACCTGCTCG 718  
Db 367 AGCAGCGGTGGCAGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 426  
Qy 719 CCGGAGAGTCTGCTTCCAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 769  
Db 427 ACGCCGAGGAGTCTGACAGCGCGCTAGACACCGCGCGCGCGCGCGCGCGCGCGCG 477

## RESULT 6

US-09-565-501A-34  
; Sequence 34, Application US/09565501A  
; Patent No. 660731  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Webb, John R.  
; APPLICANT: Dillion, Davin C.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Bhatia, Ajay  
; APPLICANT: Coler, Rhea  
; APPLICANT: Peter Probst  
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE  
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS  
; FILE REFERENCE: 210121.420C6  
; CURRENT APPLICATION NUMBER: US/09/565,501A  
; CURRENT FILING DATE: 2000-05-05  
; NUMBER OF SEQ ID NOS: 112  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 34  
; LENGTH: 516  
; TYPE: DNA  
; ORGANISM: Leishmania major  
US-09-565-501A-34

Query Match 4.4%; Score 51.4; DB 4; Length 516;  
Best Local Similarity 47.1%; Pred. No. 0.036;  
Matches 222; Conservative 0; Mismatches 246; Indels 3; Gaps 2;

Qy 302 AGTGACCGGCGAGGTCAACCGAGCTGGGTCAACGCTGGAGCGCGGCTGCGCGCGCT 361  
Db 7 AGCAGCGCTAGACACCGCCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 66  
Qy 362 GCGCGAGTCCAGCGGGTCTGCGAAGCCCTGCGAGCGAA-GACGATCAGCAATGCG 420  
Db 67 TGCCCGGACCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 126  
Qy 421 CACGGCTCTGTCACACGATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
Db 127 AGCGCTTAGACACCGCCAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 186  
Qy 481 CCGGCTCTTTCAGATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540  
Db 187 CCGCGACCGCGACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 246  
Qy 541 CCGGAGATCGGTGGCTTATCAGCGCGCTTCCCGCGCGCACTGGCGACCGGTCTCATGCTG 600  
Db 247 GCTAGACACCGCGACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 306  
Db 127 AGCGCTTAGACACCGCCAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 186

Qy 481 CCGTGTCTTTCAGATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540  
Db 187 CCGCGACCGCGACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 246  
Qy 541 CCGGAGATCGGTGGCTTATCAGCGCGCTTCCCGCGCGCACTGGCGACCGGTCTCATGCTG 600  
Db 247 GCTAGACACCGCGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 306  
Qy 601 CTGGTGGCGACCGGTCTGAGGTGGGTGAGGCGAT--CGGCTGCGCGCGCGCGCGCGCG 558  
Db 307 CGAAGCGCGAGGAGCTGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 366  
Qy 659 ACCTGCTCG 718  
Db 367 AGCAGCGGTGGCAGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 426  
Qy 719 CCGGAGAGTCTGCTTCCAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 769  
Db 427 ACGCCGAGGAGTCTGACAGCGCGCTAGACACCGCGCGCGCGCGCGCGCGCGCGCG 477

## RESULT 7

US-09-639-206A-34  
; Sequence 34, Application US/09639206A  
; Patent No. 661337  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Webb, John R.  
; APPLICANT: Dillion, Davin C.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Bhatia, Ajay  
; APPLICANT: Coler, Rhea  
; APPLICANT: Probst, Peter  
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE  
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS  
; FILE REFERENCE: 210121.420C7  
; CURRENT APPLICATION NUMBER: US/09/639,206A  
; CURRENT FILING DATE: 2000-08-14  
; NUMBER OF SEQ ID NOS: 112  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 34  
; LENGTH: 516  
; TYPE: DNA  
; ORGANISM: Leishmania major  
US-09-639-206A-34

Query Match 4.4%; Score 51.4; DB 4; Length 516;  
Best Local Similarity 47.1%; Pred. No. 0.036;  
Matches 222; Conservative 0; Mismatches 246; Indels 3; Gaps 2;

Qy 302 AGTGACCGGCGAGGTCAACCGAGCTGGGTCAACGCTGGAGCGCGGCTGCGCGCGCT 361  
Db 7 AGCAGCGCTAGACACCGCCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 66  
Qy 362 GCGCGAGTCCAGCGGGGTCTGCGAAGCCCTGCGAGCGAA-GACGATCAGCAATGCG 420  
Db 67 TGCCCGGACCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 126  
Qy 421 CACGGCTCTGTCACACGATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
Db 127 AGCGCTTAGACACCGCCAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 186  
Qy 481 CCGTGTCTTTCAGATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540  
Db 187 CCGCGACCGCGACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 246  
Qy 541 CCGGAGATCGGTGGCTTATCAGCGCGCTTCCCGCGCGCACTGGCGACCGGTCTCATGCTG 600  
Db 247 GCTAGACACCGCGACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 306  
Qy 601 CTGGTGGCGACCGGTCTGAGGTGGGTGAGGCGAT--CGGCTGCGCGCGCGCGCGCG 558

Db 307 CGAACGCCGAGAGTGTGACAGCGCCCTAGACACCGCCAGCAGCGCCGAGCTGG 366  
QY 659 ACTGCTCGCCGCGCGCCGCTGACCGTCTGTGAGCAGCTCCAGGAGTGGCCAGCA 718  
Db 367 AGGCACGGTGTGACCGCTGGCCGCGGACCGGAGCGCGCCAGCAGCTGGCCGGA 426  
QY 719 CGGAGAGTCTCTTCCAGTCCCGAGACACCGCGAAGCGCGCGCCACGG 769  
Db 427 ACCCGAGGAGTGTGACGAGCGCTAGACACCGCCAGCAGCGCGCCG 477

RESULT 8  
US-09-874-923-34  
; Sequence 34, Application US/09874923  
; Patent No. 6638517  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Webb, John R.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Bhatia, Ajay  
; APPLICANT: Coler, Rhea  
; APPLICANT: Probst, Peter  
; APPLICANT: Brannon, Mark  
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE  
; FILE REFERENCE: 210121.420C8  
; CURRENT APPLICATION NUMBER: US/09/874,923  
; CURRENT FILING DATE: 2001-06-04  
; NUMBER OF SEQ ID NOS: 122  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 34  
; LENGTH: 516  
; TYPE: DNA  
; ORGANISM: Leishmania major  
US-09-874-923-34

Query Match 4.4%; Score 51.4; DB 4; Length 516;  
Best Local Similarity 47.1%; Pred. No. 0.036;  
Matches 222; Conservative 0; Mismatches 246; Indels 3; Gaps 2;

QY 302 AGCTGAGCGGAGGTCAACAGCAGTGGTCAACAGCTGGAGCGCGCGCTGGCCCGT 361  
Db 7 AGCAGCGCTAGACACCGCCACGAGCAGCGCGCGAGCTGGAGGACCGGTGGCACGCG 66  
QY 362 GSCCGAGTCCAGCGGGGTCTCGAAGCCGCTGCGACGAA-GACGATCAGCACTGC 420  
Db 67 TGGCCGCGGACCGGACGAGCGCGCCAGCAGTGGCGCGAACCAGCGAGGAGCTGCAGC 126  
QY 421 CACGGCTCTGTGACAGATCTGCGCGCGCGGATCGCGCGAAGAGATGAGTCAAC 480  
Db 127 AGCGCTAGACACCGCCACGAGCAGCGCGCGAGCTGGAGGACCGGTGGCACGCTGG 186  
QY 481 CGTGCTCTTTCAGCAGTGTGCGCGCGCGCGCGAAGAGATGAGTCTCTGAGCGAC 540  
Db 187 CGCGGACGCGGACGAGCGCGCGCGCGCGCGAAGAGATGAGTCTCTGAGCGAC 246  
QY 541 CGGAGATCGTGGCTTATCAGCGCGCTTCCCGCGCACTGCGCGCGCGCTGATGCTG 600  
Db 247 GCTTAGACACCGCCACGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 306  
QY 601 CTGTTGGACCGGTCTGAGGTGGGTGAGGCGAT--CGGCTTGGCGCGCGCGCGCGTGG 658  
Db 307 CGAACCGGAGGAGTGTGACAGCGCTAGACACCGCCAGCAGCGCGCGCGAGCTGG 366  
QY 659 ACTGCTCGCCGCGCGCGCGCGCTGACCGTCTGTGAGCAGCTCCAGGAGTGGCCAGCA 718  
Db 367 AGGCACGGTGTGACCGCTGGCCGCGGACCGCGACGAGCGCGCGCGAGCTGGCCGGA 426  
QY 719 CGGAGAGTCTCTTCCAGTCCCGAAGACCGCGAAGCGCGCGCGCGCGCGCG 769

Db 427 AGCCCGAGGAGTGTGACAGCGCTAGACACCGCCAGCAGCGCGCGCG 477

RESULT 9  
US-09-874-923-114  
; Sequence 114, Application US/09874923  
; Patent No. 6638517  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Webb, John R.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Bhatia, Ajay  
; APPLICANT: Coler, Rhea  
; APPLICANT: Probst, Peter  
; APPLICANT: Brannon, Mark  
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE  
; FILE REFERENCE: 210121.420C8  
; CURRENT APPLICATION NUMBER: US/09/874,923  
; CURRENT FILING DATE: 2001-06-04  
; NUMBER OF SEQ ID NOS: 122  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 114  
; LENGTH: 582  
; TYPE: DNA  
; ORGANISM: Leishmania major and chagasi  
US-09-874-923-114

Query Match 4.4%; Score 51.4; DB 4; Length 582;  
Best Local Similarity 47.1%; Pred. No. 0.036;  
Matches 222; Conservative 0; Mismatches 246; Indels 3; Gaps 2;

QY 302 AGCTGAGCGGAGGTCAACAGCAGTGGTCAACAGCTGGAGCGCGCGCTGGCCCGT 361  
Db 59 AGCAGCGCTAGACACCGCCACGAGCAGCGCGCGAGCTGGCGCGAAGAGTGGCAGC 118  
QY 362 GSCCGAGTCCAGCGGGGTCTCGAAGCCGCTGCGACGAA-GACGATCAGCACTGC 420  
Db 119 TGGCCGCGGACCGGACGAGCGCGCGCGAGCTGGCGCGAAGAGTGGCAGC 178  
QY 421 CACGGCTCTGTGACAGATCTGCGCGCGCGCGATCGCGCGAAGAGTGGCAGC 480  
Db 179 AGCCCTTAGACACCGCCACGAGCAGCGCGCGAGCTGGAGGACCGGTGGCACGCTGG 238  
QY 481 CCCTGCTCTTTCAGCAGTGTGCGCGCGCGCGAAGAGATGAGTCTCTGAGCGAC 540  
Db 239 CGCGGAGCGGACGAGCGCGCGCGCGCGAGCTGGCGCGAAGAGTGGCAGC 298  
QY 541 CGGAGATCGTGGCTTATCAGCGCGCTTCCCGCGCACTGCGCGCGCGCTGCTGCTG 600  
Db 299 GCCTAGACACCGCCACGAGCAGCGCGCGAGCTGGAGGACAGGTGGCACGCTGGCCG 358  
QY 601 CTGTTGGCGCGCGCTGAGTGGGTGAGGCGAT--CGGCTTGGCGCGCGCGCGCTGG 658  
Db 359 CGAACCGGAGGAGTGTGACAGCGCTAGACACCGCCAGCAGCGCGCGCGAGTGG 418  
QY 659 ACCTGCTCGCGCGCGCGCGCGCTGACCGTCTGTGAGCAGCTCCAGGAGTGGCCAGCA 718  
Db 419 AGGCACGGTGGCAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 478  
QY 719 CGGAGAGTCTCTTCCAGTCCCGAAGACCGCGAAGCGCGCGCGCGCGCGCG 769  
Db 479 AGCCCGAGGAGTGTGACAGCGCTAGACACCGCCAGCAGCGCGCGCGCGCG 529

RESULT 10  
US-09-874-923-115  
; Sequence 115, Application US/09874923  
; Patent No. 6638517  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.

APPLICANT: Campos-Neto, Antonio  
APPLICANT: Webb, John R.  
APPLICANT: Dillon, Davin C.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Bhatia, Ajay  
APPLICANT: Coler, Rhea  
APPLICANT: Probst, Peter  
APPLICANT: Brannon, Mark  
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE  
THERAPY AND DIAGNOSIS OF LEISHMANIASIS  
FILE REFERENCE: 210121.420C8  
CURRENT APPLICATION NUMBER: US/09/874,923  
CURRENT FILING DATE: 2001-06-04  
NUMBER OF SEQ ID NOS: 122  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 115  
LENGTH: 7065  
TYPE: DNA  
ORGANISM: Leishmania major and chagasi  
US-09-874-923-115

Query Match 4.4%; Score 51.4; DB 4; Length 7065;  
Best Local Similarity 47.1%; Pred. No. 0.048;  
Matches 222; Conservative 0; Mismatches 246; Indels 3; Gaps 2;

QY 302 AGCTGACGGGAGGTACCCAGCAGTGGTCAACGACCTGGAGCGCGCTCGGCCGT 361  
DB 1283 AGCAGCGCTTAGACACCGCCACGACGCGCGCGAGCTGGAGGACCGGTGGCAGCGC 1342  
QY 362 GCGCGAGTCCACCGCGGTGCTGCGAAGCGCTGCGACGAA-GACGATCAGCAATGC 420  
DB 1343 TGGCGCGGACCGGACGAGCGCGCCAGCAGCTGCGCGGACCGGAGGAGTGCAGC 1402  
QY 421 CACGGCTGTGCACAGATCTCGCGCGGGGATCGCGCGAAGCGATCAGGCTCAAC 480  
DB 1403 AGCGCTTAGACACCGCCACGACGACGCGCGCGAGCTGGAGGACCGGTGGCAGCGCTGG 1462  
QY 481 CCGTGTCTTCGACGATGTGCCCCGCGAGCGCGAAGAGATGAAGTTCCTGAGCGAC 540  
DB 1463 CCGGACGCGGACGAGGCGCGCCACGACGCTGCGCGCGACCGCGGAGTGCAGCAGC 1522  
QY 541 CCGAGATCGTGGCTTATCAGCGCGCTTCGCGCGCATGCGCGACCGCTGCTCATGCTG 600  
DB 1523 GCCTAGACACCGCCACGACGCGCGCGAGCTGGAGGACACAGTGGCAGCGCTGGCGC 1582  
QY 601 CTGGTGGACCGGTCTGAGGTGGGTGAGGCGAT--CGGCTGCGCGCGCGCGGTGC 658  
DB 1583 CGAACCGGAGGAGTGTGAGCGCTTACCGTGTGTCAGCAGCTCCAGGAGCTGGCGCAGCA 1642  
QY 659 ACCTGTGCGCGCGCGCGCGCTGACCGTGTGTCAGCAGCTCCAGGAGCTGGCGCAGCA 718  
DB 1643 AGSCACGGGTGGCACGCTGCGCGCGACCGCGAGCGCGCGCAGCAGCTGGCGCGCA 1702  
QY 719 CGGAGAGCTGCTTCCAGTCCCGACGACCGCGAGGCGCGCGCGCG 769  
DB 1703 ACGCCGAGGAGTGTGACGAGCGCTTAGACACCGCCACGACGAGCGCGCG 1753

RESULT 11  
US-09-679-279-1  
Sequence 1, Application US/09679279  
Patent No. 6524841  
GENERAL INFORMATION:  
APPLICANT: McDaniel, Robert  
APPLICANT: Volchegursky, Yanina  
TITLE OF INVENTION: Recombinant Megalomicin Biosynthetic  
TITLE OF INVENTION: Genes and Uses Thereof  
FILE REFERENCE: 300622004700  
CURRENT APPLICATION NUMBER: US/09/679,279  
CURRENT FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/158,305  
PRIOR FILING DATE: 1999-10-08  
PRIOR APPLICATION NUMBER: US 60/190,024

PRIOR FILING DATE: 2000-03-17  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 47981  
TYPE: DNA  
ORGANISM: Micromonospora megalomicea  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(144)  
OTHER INFORMATION: megBVI (megT), TDP-4-keto-6-deoxyglucose-2,3-dehydratase;  
OTHER INFORMATION: SEQ ID NO: 2= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (928)...(2061)  
OTHER INFORMATION: megDVI, TDP-4-keto-6-deoxyglucose 3,4-isomerase,  
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,4-isomerase;  
OTHER INFORMATION: SEQ ID NO: 3= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (2072)...(3382)  
OTHER INFORMATION: rhodosaminyl transferase (eryCIII homolog),  
OTHER INFORMATION: TDP-megosamine glycosyltransferase;  
OTHER INFORMATION: SEQ ID NO: 4= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (3462)...(4634)  
OTHER INFORMATION: megG (megY), mycetosyl acyltransferase, mycotoxose O-acyltransferase  
OTHER INFORMATION: SEQ ID NO: 5= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (4651)...(5775)  
OTHER INFORMATION: megDII, deoxysugar transaminase (eryCI, DnrJ homolog),  
OTHER INFORMATION: TDP-3-keto-6-deoxyhexose 3-aminotransaminase;  
OTHER INFORMATION: SEQ ID NO: 6= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (5822)...(6595)  
OTHER INFORMATION: megDIII, daunosaminyl-N,N-dimethyltransferase (eryCVI homolog);  
OTHER INFORMATION: SEQ ID NO: 7= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (6592)...(7197)  
OTHER INFORMATION: megDIV, TDP-4-keto-6-deoxyglucose 3,5-epimerase (eryBVII, dmU hor  
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,5-epimerase;  
OTHER INFORMATION: SEQ ID NO: 8= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (7220)...(8206)  
OTHER INFORMATION: megDV, TDP-hexose 4-ketoreductase (eryBIV, dmV homolog),  
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;  
OTHER INFORMATION: SEQ ID NO: 9= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (8228)...(9220)  
OTHER INFORMATION: megEII-1 (megDVII), TDP-4-keto-L-6-deoxy-hexose 2,3-reductase;  
OTHER INFORMATION: SEQ ID NO: 10= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (9226)...(10479)  
OTHER INFORMATION: megBV, mycetosyl transferase, mycotoxose glycosyltransferase;  
OTHER INFORMATION: SEQ ID NO: 11= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (10483)...(11424)  
OTHER INFORMATION: megBIV, TDP-hexose 4-ketoreductase,  
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;  
OTHER INFORMATION: SEQ ID NO: 12= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (12181)...(22821)  
OTHER INFORMATION: megAI; SEQ ID NO: 13= translated amino acid sequence  
NAME/KEY: misc feature  
LOCATION: (12505)...(13470)  
OTHER INFORMATION: megAI, AT-L  
NAME/KEY: misc feature  
LOCATION: (13576)...(13791)  
OTHER INFORMATION: megAI, ACP-L  
NAME/KEY: misc feature  
LOCATION: (13849)...(15126)  
OTHER INFORMATION: megAI, KS1  
NAME/KEY: misc feature  
LOCATION: (15427)...(16476)  
OTHER INFORMATION: megAI, AT1

```
; NAME/KEY: misc_feature
; LOCATION: (17155)...(17694)
; OTHER INFORMATION: megAI, KR1
; NAME/KEY: misc_feature
; LOCATION: (17947)...(18207)
; OTHER INFORMATION: megAI, ACP1
; NAME/KEY: misc_feature
; LOCATION: (18268)...(19548)
; OTHER INFORMATION: megAI, KS2
; NAME/KEY: misc_feature
; LOCATION: (19876)...(20910)
; OTHER INFORMATION: megAI, AT2
; NAME/KEY: misc_feature
; LOCATION: (21517)...(22053)
; OTHER INFORMATION: megAI, KR2
; NAME/KEY: misc_feature
; LOCATION: (22318)...(22575)
; OTHER INFORMATION: megAI, ACP2
; NAME/KEY: CDS
; LOCATION: (22867)...(33555)
; OTHER INFORMATION: megAI; SEQ ID NO: 14= translated amino acid sequence
; NAME/KEY: misc_feature
; LOCATION: (22957)...(24237)
; OTHER INFORMATION: megAI, KS3
; NAME/KEY: misc_feature
; LOCATION: (24544)...(25581)
; OTHER INFORMATION: megAI, AT3
; NAME/KEY: misc_feature
; LOCATION: (26230)...(26733)
; OTHER INFORMATION: megAI, KR3 (inactive)
; NAME/KEY: misc_feature
; LOCATION: (26938)...(27258)
; OTHER INFORMATION: megAI, ACP3
; NAME/KEY: misc_feature
; LOCATION: (27393)...(28590)
; OTHER INFORMATION: megAI, KS4
; NAME/KEY: misc_feature
; LOCATION: (28897)...(29931)
; OTHER INFORMATION: megAI, AT4
; NAME/KEY: misc_feature
; LOCATION: (29953)...(30477)
; OTHER INFORMATION: megAI, DH4
; NAME/KEY: misc_feature
; LOCATION: (31396)...(32244)
; OTHER INFORMATION: megAI, ER4
; NAME/KEY: misc_feature
; LOCATION: (32257)...(32799)
; OTHER INFORMATION: megAI, KR4
; NAME/KEY: misc_feature
; LOCATION: (33052)...(33312)
; OTHER INFORMATION: megAI, ACP4
; NAME/KEY: CDS
; LOCATION: (33666)...(43271)
; OTHER INFORMATION: megAI; SEQ ID NO: 15= translated amino acid sequence
; NAME/KEY: misc_feature
; LOCATION: (33780)...(35027)
; OTHER INFORMATION: megAI, KS5
; NAME/KEY: misc_feature
; LOCATION: (35385)...(36419)
; OTHER INFORMATION: megAI, AT5
; NAME/KEY: misc_feature
; LOCATION: (37068)...(37604)
; OTHER INFORMATION: megAI, KR5
; NAME/KEY: misc_feature
; LOCATION: (37860)...(38120)
; OTHER INFORMATION: megAI, ACP5
; NAME/KEY: misc_feature
; LOCATION: (38187)...(39470)
; OTHER INFORMATION: megAI, KS6
; NAME/KEY: misc_feature
; LOCATION: (39795)...(40811)
; OTHER INFORMATION: megAI, AT6
; NAME/KEY: misc_feature
```

```
; LOCATION: (41406)...(41936)
; OTHER INFORMATION: megAI, KR6
; NAME/KEY: misc_feature
; LOCATION: (42168)...(42425)
; OTHER INFORMATION: megAI, ACP6
; NAME/KEY: misc_feature
; LOCATION: (42585)...(43271)
; OTHER INFORMATION: megAI, TE
; NAME/KEY: CDS
; LOCATION: (43268)...(44344)
; OTHER INFORMATION: megAI, TDP-4-keto-6-deoxyglucose 3,4-isomerase;
; OTHER INFORMATION: SEQ ID NO: 16= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (44355)...(45623)
; OTHER INFORMATION: megAI, desosaminyl transferase, desosamine glycosyltransferase;
; OTHER INFORMATION: SEQ ID NO: 17= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (45620)...(46591)
; OTHER INFORMATION: megAI-2 (megBI), TDP-4-keto-6-deoxy-L-glucose 2,3 dehydratase,
; OTHER INFORMATION: TDP-4-keto-6-deoxyglucose 2,3 dehydratase;
; OTHER INFORMATION: SEQ ID NO: 18= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (46660)...(47403)
; OTHER INFORMATION: megH, TEII; SEQ ID NO: 19= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (47411)...(47980)
; OTHER INFORMATION: megP, C-6 hydroxylase; SEQ ID NO: 20= translated amino acid sequence
; US-09-679-279-1

Query Match      4.3%; Score 50.4; DB 4; Length 47981;
Best Local Similarity 46.3%; Pred. No. 0.096;
Matches 251; Conservative 0; Mismatches 276; Indels 15; Gaps 2;

QY 219 GCTGAACCGACGCGCGTGAACTCGGAGGCAACCGGATCCGCAACACCATCTCTGCCCAT 278
DB 29721 GCTGATGGCGTTCGAGAGACCGCCGAGAGCGCGGCGGCTCACCGCGTCCGAC 29780
QY 279 ACTCGGCATCTACCTTTGACGAGCTGAGCGGAGGTACCCAGAGTGGTTCACGA 338
DB 29781 GCTGCGCGGACCAACGACGCGGCGTCTCCGCAACCTCTCGGCGGCGACGT 29840
QY 339 CCTGAGGCGCGGTCGCGCGGTCGCGGAGTCCACGCGGGTCTCGGAAGCGCTGGC 398
DB 29841 GCACGGGTTCAGTTCAGTTCGCGGCGTCCGCCAGCGCGCTGGTTCGACCTGCC 29900
QY 399 AGCGAAGACGATCAGCAACTGCCACGCGCTGTGTCACACGATCTGCGCGGCGGATCGC 458
DB 29901 CACCTACCCCTTCGACAGGCGGCTCTGCGCCCAAGCCGACCGCAGGCGCGACCTC 29960
QY 459 GCGCAAGCGATCAGGCTCAACCGCTCTTCGACGATGCTGCCCGCGCGAGCGAA 518
DB 29961 GTCGCTGGGGTCCGTGAC-----TCGACCCACCGCTGTGTGACCGCGAGT 30008
QY 519 AGAGATGAAGTTCCTGAGCGACCCGAGATCGGTTCGGCTTATACAGCGCGTTCCGCCGA 578
DB 30009 CGACGTACCCGGTCAACGCGGAGCGGTTTCACCGGCGGCTCTCCCGCAGCAGCA 30068
QY 579 CTGGCGACCGCTCTGATGTGTGTGGGACCGGTCTGAGTGGGTGGGTGAGCGATCGG 638
DB 30069 GTGGCTGACCCAGCACGTGTGGTGGGCGCAACTGTCGCCGCGAGTTCCTGGTCA 30128
QY 639 CCT---GCGCGCGCGGGTTCGACCTGCTCGCGCGCGCGCGCGCTGACCGTCTCGA 695
DB 30129 CCTCGCGCTACCGCGGCGCGCGAGCTCGGCGTCCCGGTCTGGAGGAACCTGCTCTCA 30188
QY 696 CGAGCTCCAGAGTGGCCAGCACGAGAGAGTCTGTTTCCAGTCCGCGGAGACCGGAA 755
DB 30189 GCAGCGGTGTTGTTGACCGCGCGGTGCTGCTGCGCTGCTGCTGCGCGCGCGCA 30248
QY 756 GG 757
DB 30249 CG 30250
```

RESULT 12  
US-09-252-991A-11386/c  
; Sequence 11386, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 11386  
; LENGTH: 534  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11386

Query Match 4.3%; Score 50.2; DB 4; Length 534;  
Best Local Similarity 46.2%; Pred. No. 0.064;  
Matches 166; Conservative 0; Mismatches 193; Indels 0; Gaps 0;  
QY 447 CGCGCGATCGCGCGAAGCGATCAGGCTCAACCCGCTCTTCGACGATGCTGCCCG 506  
DB 463 CGCGGAGAGCGGATCGGAGACCGCCAGGCTCACCGGAACTGGCGGTAGATGTTGCCAC 404  
QY 507 GCCTGCGGCGACGATGAGTTCCTGAGCGACCGCGAGATCGGTCTTATCACCG 566  
DB 403 CGCGCGCTGAGAGACCGCATCGGACGAAACACCGAGACGACACCGGTGATGCCGAC 344  
QY 567 GCTTCGCGCGCACTGGCGACCGCTCGTCATGCTGTGGGACCGGCTCTGAGGTGGG 626  
DB 343 GATGGCGCGCTGATCGGCGCATCGCTTGACCGTGGGTCTGGCGGACAGGCTTC 284  
QY 627 TGAGCGATCGGCTGGCGCGCGCGCGGCTGACCTGCTTCCAGTCCCGGAA 746  
DB 283 CTGCGGCATCAGCGCTGCGACGCTTCCACGAGTATCGGTCTGTCACGAGATGCC 224  
QY 687 CGTCTGCGACGCTCGAGAGCTGCGGACGCGGAGAGCTCGTCTTCCAGTCCCGAA 746  
DB 223 GATCGCGCAGGACCATCGGAAACATGTCAGCAGTTGATCGAGAGCCAGCGCGAT 164  
QY 747 GACCGGAGGCGCGCGCGCGCGCGCTGAGTTTACCGAGAAAGTCTTACTCTTACGC 805  
DB 163 CACCGTGAAGGTGCCCGAGCGGCGGACCGGCGACCGGCGGACCGTCTGCGGATCAGG 105

RESULT 13  
US-09-252-991A-11487  
; Sequence 11487, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 11487  
; LENGTH: 3150  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11487

Query Match 4.3%; Score 50.2; DB 4; Length 3150;  
Best Local Similarity 46.2%; Pred. No. 0.078;  
Matches 166; Conservative 0; Mismatches 193; Indels 0; Gaps 0;  
QY 447 CGCGCGATCGCGCGAAGCGATCAGGCTCAACCCGCTCTTCGACGATGCTGCCCG 506  
DB 1702 CGCGGAGAGCGGATGAGAGCGGCGGATCAGCGGAACTGGCGGTAGATGTTGCCAC 1761  
QY 507 GCCTGCGGCGAAGAGATGAGTTCCTGAGCGACCGCGAGATCGGTGGCTTATCACGGC 566  
DB 1762 CGCGCGCTGAGAGACCGCATCGGACGAAACACCGAGACCGACGAGTGTGCCGAC 1821  
QY 567 GCTTCGCGCGCACTGGCGACCGCTCGTCATGCTGTGGCGACCGGCTGAGGTGGG 626  
DB 1822 GATGGCGCGCTGATCTGGCGCATCGGCTTGACCGGTGGCTGTCGCGGACAGGCTTC 1881  
QY 627 TGAGCGATCGGCTGGCGCGCGCGGCTGACCTGCTCGCGCGCGCGCGGCTGAC 686  
DB 1882 CTGCGCATCAGCGCTGCGGCTTCTCCACGATGATCGGTGTCACCGAGTGC 1941  
QY 687 CGTCTGCGACGATCCAGGAGCTGGCGGACGCGGAGAGCTCGTCTTCCAGTCCCGAA 746  
DB 1942 GATCGCGAGGACCATCGGAAACATGTCAGCAGTTGATCGAGAGCCAGGCGGAT 2001  
QY 747 GACCGGAGGCGCGCGCGCGCGGTCAGTTCACCGAGAAAGTCTTACTGTTAGCG 805  
DB 2002 CACCGTGAAGGTGCCCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2060

RESULT 14  
US-09-252-991A-11405/c  
; Sequence 11405, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 11405  
; LENGTH: 3372  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11405

Query Match 4.3%; Score 50.2; DB 4; Length 3372;  
Best Local Similarity 46.2%; Pred. No. 0.078;  
Matches 166; Conservative 0; Mismatches 193; Indels 0; Gaps 0;  
QY 447 CGCGCGATCGCGCGAAGCGATCAGGCTCAACCCGCTCTTCGACGATGCTGCCCG 506  
DB 1458 CGCGGAGAGCGGATGAGAGCGGCGGATCAGCGGAACTGGCGGTAGATGTTGCCAC 1399  
QY 507 GCCTGCGGCGAAGAGATGAGTTCCTGAGCGACCGCGGAGATCGGTGGCTTATCACGGC 566  
DB 1398 CGCGCGCTGAGAGACCGCATCGGACGAAACACCGAGACCGACCGGTGATGCCGAC 1339  
QY 567 GCTTCGCGCGCACTGGCGACCGCTCGTCATGCTGTGGCGACCGGCTGAGGTGGG 626  
DB 1338 GATGGCGCGCTGATCTGGCGCATCGCTTGACCGTGGGTCTGCGGCGACAGGCTTC 1279  
QY 627 TGAGCGATCGGCTGGCGCGCGCGGCTGACCTGCTGCGCGCGCGGCGGCGGCTGAC 686  
DB 1278 CTGCGCATCAGCGCTGCGGCTTCTCCACGATGATCGGTCTGTCACCGAGTGC 1219  
QY 687 CGTCTGCGACGATCCAGGAGCTGGCGGAGAGCTCGTCTTCCAGTCCCGGAA 746

Db 1218 GATCCGAGGACCATGCCGACATGCTCAGACAGTGTGATCAGAGGCCAGCCGAGCAT 1159  
Qy 747 GACCGGAGGCGCGGCGACGCTCAGTTTACACGAAAGTCGCTTACTGCTTACGC 805  
Db 1158 CACCGTGAAGGTGCCGAGCGGCCACCGGACCAACCACTGTCGGGATCAGGCTGGGCG 1100

RESULT 15  
US-09-252-991A-2186  
; Sequence 2186, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 2186  
; LENGTH: 915  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-2186

Query Match 4.1%; Score 48.8; DB 4; Length 915;  
Best Local Similarity 51.1%; Pred. No. 0.13;  
Matches 140; Conservative 0; Mismatches 132; Indels 2; Gaps 1;  
Qy 178 GCGATTTCGTGGGAGTGTGGCGGAGCTACGAAAGACGCTGAAACCGACGCC--G 235  
Db 395 GCCAGTTGCCGATGGACGTGGTGGCGGTGATTTCCAAATCACCAGCCTGGAGCGCTGG 454  
Qy 236 TGAATCTCGAGGGGCAACCGGATCCGCAACCACTCTCTGCCATCTCGGCCATCTCACCC 295  
Db 455 CGCGCTGGCAGCGCATTCCTACCACCACTTCCCGCTCGACCCGAAACGACAGCGCGGCC 514  
Qy 296 TTGACGAGCTGACGGGAGGTCAACGACAGTGGGTCAACGACCTGGAGCGCGGCTCG 355  
Db 515 AGGAAGCCAGGCTCTGGCAGGTGCTCGAAGAGAGTGGCGCGCGAGCTGGTGATCCTCGCTC 574  
Qy 356 GCCGTGGCGGAGTCCACGCGGGTCTCGGAAGCGGCTGGCAGCGAAGACGATCAGCA 415  
Db 575 GTACATGAGGTGCTGTGCGGGAATCTGCGGGGCTGGACGGCTGGGGGATCAACA 634  
Qy 416 ATGCCACGGCTGTGACACAGATCTCGGGGC 449  
Db 635 TCCATCACTCCCTGCTGCCCGGCTTCAGGGGCGC 668

Search completed: August 6, 2004, 13:42:24  
Job time : 147.71 secs

**This Page Blank (uspto)**

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2004, 08:50:04 ; Search time 1956.12 Seconds  
(without alignments)  
9439.175 Million cell updates/sec

Title: US-09-855-340A-2

Perfect score: 426

Sequence: 1 atcgcaacacacacggggct.....acctgtccgcggagctga 426

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:  
1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_ov.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_or.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vl.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pln.\*  
35: em\_htg\_rod.\*  
36: em\_htg\_mam.\*  
37: em\_htg\_vrt.\*  
38: em\_sy.\*  
39: em\_htgo\_hum.\*  
40: em\_htgo\_mus.\*  
41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	426	100.0	426	6	AX338971 Sequence
2	426	100.0	2025	7	AY150027 Bacteriopl
3	66	15.5	1575	6	E03045 DNA encodin
4	64.6	15.2	1653	8	AK121255 Oryza sat
5	64.6	15.2	124629	8	OSU000082 Oryza sat
6	64.4	15.1	1056	6	BD269437 Mitomycin
7	64.4	15.1	1056	6	AR266923 Sequence
8	64.4	15.1	53500	6	BD269447 Mitomycin
9	64.4	15.1	53500	6	AR266933 Sequence
10	64.4	15.1	63734	1	AF127374 Streptomy
11	64.4	15.1	138203	1	AY310323 Streptomy
12	63.4	14.9	145539	8	CNS08CEV Oryza sat
13	62.6	14.7	276800	1	SC0939115 Streptomy
14	61.6	14.5	2081	1	SSBGL3GE Streptomy
15	61.4	14.4	1233	6	BD179992 Highly th
16	61.2	14.4	10352	1	AF546152 Streptomy
17	61	14.3	110000	2	AC139251 Continuation (2 of
18	60.6	14.2	77457	1	AF210249 Streptomy
19	59.6	14.0	132544	1	AF521085 Streptomy
20	59	13.8	3198	1	AF233851 Streptomy
21	58.6	13.8	296300	1	AP005035 Streptomy
22	58.6	13.8	300425	1	AP005038 Streptomy
23	58.2	13.7	125020	9	AF429315 Homo sapi
24	58.2	13.7	349315	1	BX572593 Rhodopseu
25	58	13.6	299425	1	AP005037 Streptomy
26	57.8	13.6	1193	8	AK109088 Oryza sat
27	57.8	13.6	143407	8	AP003825 Oryza sat
28	57.8	13.6	299925	1	AP005045 Streptomy
29	57.6	13.5	120627	2	AC104714 Oryza sat
30	57.4	13.5	18209	1	AE005129 Halobacte
31	57.4	13.5	197050	1	AL646081 Ralstonia
32	57.4	13.5	321250	1	SC0939111 Streptomy
33	57.2	13.4	292100	1	SC0939121 Streptomy
34	57	13.4	15738	6	AX803762 Sequence
35	57	13.4	61944	6	AX803750 Sequence
36	56.8	13.3	972	6	AX574166 Sequence
37	56.8	13.3	1195	8	AK110593 Oryza sat
38	56.8	13.3	45055	6	AX574197 Sequence
39	56.8	13.3	138390	14	AY261359 Bovine he
40	56.6	13.3	303550	1	SC0939131 Streptomy
41	56.4	13.2	1381	14	AY299389 Ori virus
42	56.4	13.2	58343	1	AB089954 Micromono
43	56.4	13.2	310550	1	SC0939113 Streptomy
44	56.2	13.2	420	1	MXU49152 Myxococcus
45	56.2	13.2	923	8	AK063822 Oryza sat

ALIGNMENTS

RESULT 1	AX338971	Sequence 2 from Patent WO0187936.	426 bp	DNA	linear	PAT 09-JAN-2002
LOCUS	AX338971					
DEFINITION	AX338971					
ACCESSION	AX338971					
VERSION	AX338971.1	GI:18129107				
KEYWORDS						
SOURCE						
ORGANISM						

Micromonospora carbonacea  
Micromonospora carbonacea  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Micromonosporineae; Micromonosporaceae; Micromonospora.  
1  
Hosted, T.J. and Horan, A.C.  
Isolation of Micromonospora carbonacea var africana pmlp1 integrase  
and use of integrating function for site-specific integration into

Micromonospora halophitica and Micromonospora carbonacea chromosome  
Patent: WO 0187936-A 2 22-NOV-2001;  
SCHERING CORPORATION (US)

## FEATURES

source  
1. .426  
/organism="Micromonospora carbonacea"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:47853"

## ORIGIN

Query Match 100.0%; Score 426; DB 6; Length 426;  
Best Local Similarity 100.0%; Pred. No. 5.4e-42;  
Matches 426; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGGGMAACACACGGGGCTGGGGCGGGGACATATGGGGCGCATACCTCTACCGCCCGC 60  
DB 1 ATGCGCAACACACGGGGCTGGGGCGGGGACATATGGGGCGCATACCTCTACCGCCCGC 60  
QY 61 GAGCGCCCGGACTGACCAAGAGCGAGTTGGCCAGGCGCATCCAGAAAGACCGGGCCACC 120  
DB 61 GAGCGCCCGGACTGACCAAGAGCGAGTTGGCCAGGCGCATCCAGAAAGACCGGGCCACC 120  
QY 121 GTGGGCGGTGGAGGACGCAAGAACCGGCCCGCAGCGCGACCTCTGTGCGCGCTC 180  
DB 121 GTGGGCGGTGGAGGACGCAAGAACCGGCCCGCAGCGCGACCTCTGTGCGCGCTC 180  
QY 181 GCCAGGTGTCGGCCTCGACCTCGACGAGCCCTCGCGCGCGAGCTCTGGCGCCCGC 240  
DB 181 GCCAGGTGTCGGCCTCGACCTCGACGAGCCCTCGCGCGCGAGCTCTGGCGCCCGC 240  
QY 241 GTCAACCGCGCAGCGACCCCAACCATGACCTGAGCGAGAAATCGAGTGTGCGCAC 300  
DB 241 GTCAACCGCGCAGCGACCCCAACCATGACCTGAGCGAGAAATCGAGTGTGCGCAC 300  
QY 301 GACCCCAAGCTGACGAGGACATGACGCGCGCATATCGCCCTATCTGGAGCGCGT 360  
DB 301 GACCCCAAGCTGACGAGGACATGACGCGCGCATATCGCCCTATCTGGAGCGCGT 360  
QY 361 GAGCGGACAAAGCGGGCGGCGATCGAGGAAACCAAGCGGCTATCGACCTGTTCGCGCG 420  
DB 361 GAGCGGACAAAGCGGGCGGCGATCGAGGAAACCAAGCGGCTATCGACCTGTTCGCGCG 420  
QY 421 AGCTGA 426  
DB 421 AGCTGA 426

## RESULT 2

AY150027  
LOCUS Bacteriophage pMLP1 att/int region. 2025 bp DNA linear PHG 02-SEP-2003  
DEFINITION Bacteriophage pMLP1 att/int region.  
ACCESSION AY150027  
VERSION AY150027.1 GI:28630421

## KEYWORDS

Bacteriophage pMLP1

## SOURCE

Bacteriophage pMLP1

## ORGANISM

Viruses.

1 (bases 1 to 2025)

Alexander, D.C., Devlin, D.J., Hewitt, D.D., Horan, A.C. and

Hosted, T.J.

Development of the Micromonospora carbonacea var. africana ATCC

39149 bacteriophage pMLP1 integrase for site-specific integration

in Micromonospora spp

Microbiology (Reading, Engl.) 149 (Pt 9), 2443-2453 (2003)

12949170

2 (bases 1 to 2025)

Hosted, T.J. Jr., Alexander, D.C. and Hewitt, D.D.

Direct Submission

Submitted (12-SEP-2002) New Lead Discovery, Schering-Plough

Research, 2015 Galloping Hill Road, Kenilworth, NJ 07033, USA

## FEATURES

source  
1. .2025  
/organism="Bacteriophage pMLP1"

/mol\_type="genomic DNA"  
/specific\_host="Micromonospora sp. ATCC 39149"  
/db\_xref="taxon:219292"  
/note="host is deposited in ATCC as Micromonospora carbonacea var. africana Waitz et al."  
misc\_feature  
1. .2025  
/note="att/int region"  
CDS  
197..622  
/note="xis"  
/codon\_start=1  
/transl\_table=11  
/product="excisionase"  
/protein\_id="AAO46044.1"  
/db\_xref="GI:28630422"  
/translation="MRNTPGLRGTHWAAVYLTARERAGLTYSLARIRIQDRAVGRW  
EDGNRPDDADLVARVAVGLDDEAAGAARPGVTPPATFMDLDEIELVRIDP  
KLDEMKRRIIALILERRERDKAAATSETRKLDLFRS"  
628..1806  
/gene="intM"  
628..1806  
/gene="intM"  
/codon\_start=1  
/transl\_table=11  
/product="integrase"  
/protein\_id="AAO46045.1"  
/db\_xref="GI:28630423"  
/translation="MWIEKNPVVIRIDLVRGKVTIOTGYPTKSAKNMVQFRAEQ  
LQNALMPRGQITLADPFGEMWPSYEKTLKPTAVNSEGNRIHLLPILGHYLTDEL  
DQVQTVWVNDLEAGVGPWESETRGRKPLAAKTI SNCHGLLHTI GGAIAAKRIRLN  
PCSTMLPRPREKEMFLSDPEIGRLITLAPPHWRFLVLLVAITGKRWBAIGRAGR  
VDLLAARPLTVVEQLELASTGELVFQSPKAGRTVSFTTKVALLLTLIAGKRS  
DEVVTAAPKGMVTRNFRIRIVKACEEAGLPGLRIHDLRHTAAI IISAGRLPISIS  
RLHSHSIAVTDLLYGLHREVDGILAAIEEAMAGVRAEDLEAEDELDTDLADAA  
"  
stem\_loop  
join(1884..1895,1900..1911)  
misc\_feature  
1925..1986  
/note="attP element"  
stem\_loop  
join(1969..1986,1991..2010)

Query Match 100.0%; Score 426; DB 7; Length 2025;  
Best Local Similarity 100.0%; Pred. No. 3.5e-42;  
Matches 426; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGCGCAACACACCGGGCTGGGGCGGGCAATGGCGCGCATACGTCCTACCGCCCGC 60  
DB 197 ATGCGCAACACACCGGGCTGGGGCGGGCAATGGCGCGCATACGTCCTACCGCCCGC 256  
QY 61 GAGCGCCCGGACTGACCAAGAGCGAGTTGGCCAGGCGCATCCAGAAAGACCGGGCCACC 120  
DB 257 GAGCGCCCGGACTGACCAAGAGCGAGTTGGCCAGGCGCATCCAGAAAGACCGGGCCACC 316  
QY 121 GTGGCGCGGTGGGAGGACGGCAAGAACCGGCCCGACGACCGGACTCTGTTGCCCGCTC 180  
DB 317 GTGGCGCGGTGGGAGGACGGCAAGAACCGGCCCGACGACCGGACTCTGTTGCCCGCTC 376  
QY 181 GCCCAGGTGTCGGCTCGACCTCGACGAAGCCCTCGCGCGCGAGTCTGCGCCCGCGC 240  
DB 377 GCCCAGGTGTCGGCTCGACCTCGACGAAGCCCTCGCGCGCGAGTCTGCGCCCGCGC 436  
QY 241 GTACCCCGCGCAGCGACCCCAACCATCGAGCTGGACGAGAAATCGAGCTGGTCCCGCAC 300  
DB 437 GTACCCCGCGCAGCGACCCCAACCATCGAGCTGGACGAGAAATCGAGCTGGTCCCGCAC 496  
QY 301 GACCCCAAGCTGACGAGGACATGAACGGCGGATCATTCGCCCTATCTCTGGAGCGCGCT 360  
DB 497 GACCCCAAGCTGACGAGGACATGAACGGCGGATCATTCGCCCTATCTCTGGAGCGCGCT 556  
QY 361 GAGCGGACAAAGCGGGCGGCGATCGAGGAAACCAAGCGGCTCATCGACCTGTTCGCGCG 420  
DB 557 GAGCGGACAAAGCGGGCGGCGATCGAGGAAACCAAGCGGCTCATCGACCTGTTCGCGCG 616  
QY 421 AGCTGA 426

```

Db          617 AGCTGA 622          |||||
RESULT 3
E03045
LOCUS      E03045          1575 bp      DNA      linear      PAT 29-SEP-1997
DEFINITION DNA encoding streptomycetes chromofuscus phosphatidyl choline
            phosphatidohydrolase.
ACCESSION E03045
VERSION   E03045.1 GI:2171267
KEYWORDS  JP 1991187382-A/1.
SOURCE    Streptomycetes chromofuscus
ORGANISM  Streptomycetes chromofuscus
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomycetes.
REFERENCE 1 (bases 1 to 1575)
AUTHORS   Horiuchi.K., Mizoguchi,J., Takahara.M., Imamura,S., Beppu,T. and
            Horiuchi,S.
TITLE     DNA CONTAINING GENETIC INFORMATION OF PHOSPHOLIPASE AND USE THEREOF
JOURNAL
COMMENT   OS Streptomycetes chromofuscus
            Patent: JP 1991187382-A 1 15-AUG-1991;
            TOYO JOZO CO LTD
            PN JP 1991187382-A/1
            PD 15-AUG-1991
            PF 15-DEC-1989 JP 1989325355
            PI HOUJIYU KAZUO, MIZOGUCHI JUNZO, TAKAHARA MASAYASU, PI
            IMAMURA SHIGEKI,
            PI BEPPU TERUHIKO, HORINOCHI SUEJI
            PC C12N15/55,C12N1/19,C12N1/21,C12N9/20,(C12N1/19,C12R1:865), PC
            (C12N1/21,
            PC C12R1:19),(C12N1/21,C12R1:465),(C12N1/21,C12R1:125); CC
            strandedness: Double;
            CC topology: Linear;
            CC hypothetical: No;
            CC anti-sense: No;
            CC *source: strain=A-0848;
            CC *source: developmental_stage=Germ-Line;
            CC *source: clone=pcPLD1;
            PH Key
            FH Location/Qualifiers
            FT mat_peptide 43..1572
            FT /product='Streptomycetes chromofuscus FT
            Phosphatidylcholine
            Phosphatidohydrolase'.
FEATURES
source
1..1575
Location/Qualifiers
/organism="Streptomycetes chromofuscus"
/mol_type="genomic DNA"
/db_xref="taxon:42881"
ORIGIN
Query Match 15.5%; Score 66; DB 6; Length 1575;
Best Local Similarity 48.4%; Pred. No. 29;
Matches 183; Conservative 0; Mismatches 195; Indels 0; Gaps 0;
QY 7 AACACACGGGGCTGGGGCGGGGCACATGGGCCCGCATACGTCCTACCGCCCGCGAGCGC 66
Db 646 AAGCGCGCGTCTGTCGGGATCTGGGAGCACCACAGATTTCGCGCAACACGCTGTGCGGC 705
QY 67 GC CGGACTGACCAAGAGCGAGTTGGCCAGGCGCATCCAGAGGACCGGGCCACCGTCGCGC 126
Db 706 GGC CGGTGAACCAACGAGGGCGCCGAGGGCACTTGTCGGCCCGTCAGGCCGCGCC 765
QY 127 CGGTGGGAGGACGGCAAGAACCCGGCCCGACGACGCGGACCTCGTTGCCCGCGTCCGCCAG 186
Db 766 AAGCAGCGCTACTTCGAGTGTGATCGCGTGGCGCCCGCCATCGCGCGCACCCACCTACCGG 825
QY 187 GTGCTCGCCTCGACCTCGAGAGGCCTCGCCCGCCGAGTCTGCGCCCGCGGTCAACC 246
Db 826 CGGCTCGCCTTCGGAAGCTCGCCGACCTCTTCCTGTGGACCTGCGCTCTTCCTCGCTCC 885
QY 247 CCGCCAGCGACCCCAACCATGACCTGGACGAGGAAATCGAGCTGTGTCGCGACCGACCCC 306

```

REFERENCE 3 (bases 1 to 1653)  
 TITLE Kikuchi,S.  
 AUTHORS Direct Submission  
 JOURNAL Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of  
 Agrobiological Sciences, Department of Molecular Genetics, Head of  
 Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki  
 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp,  
 tel:81-29-839-7007, Fax:81-29-838-7007)  
 COMMENT This clone is one of the 32K full-length cDNA clones from japonica  
 rice.  
 URL : http://cdna01.dna.affrc.go.jp/cDNA/  
 NTAS Rice Full-length cDNA Project Team: Kikuchi,S., Satch,K.,  
 Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J.,  
 Ishikawa,M., Yanada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T.,  
 Onoda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T.,  
 Yamamoto,M. and Nakahama,Y.  
 FAIS Genome Sequencing & Analysis Group: Ootomo,Y., Iida,Y.,  
 Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M.,  
 Kobayashi,T., Kodama,T., Kurosaki,T., Kusunagi,T., Lu,M.,  
 Masuda,H., Miura,J., Mizuno,K., Narikawa,R., Niikura,J., Oka,M.,  
 Ryu,R., Sugano,S., Sugiyama,A., Suzuki,Y., Tsunoda,Y., Teda,M.,  
 Xie,Q., Yokomizo,S., Yoshimura,A., Matsubara,K. and Murakami,K.  
 Genome Exploration Research Group in Riken Genomic Sciences Center  
 and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K.,  
 Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T.,  
 Hara,A., Hashizume,M., Hayaishi,K., Hayatsu,N., Hiramoto,K.,  
 Hiraoka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y.,  
 Itoh,M., Kagawa,I., Kanagawa,S., Kato,H., Kawai,J.,  
 Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Komoto,H., Kouda,M.,  
 Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,  
 Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N.,  
 Oca,Y., Saichoh,H., Sakai,C., Sakai,K.,  
 Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K.,  
 Shingawa,A., Shiraki,T., Sogabe,Y., Tagami,M.,  
 Tagami-Takeda,Y., Tagawa,A., Takahashi,F.,  
 Takaku-Akahara,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K.,  
 Yasunishi,A. and Hayashizaki,Y.  
 Location/Qualifiers  
 FEATURES  
 source  
 1..1653  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="J023102B22"  
 ORIGIN  
 Query Match 15.2% Score 64.6; DB 8; Length 1653;  
 Best Local Similarity 49.6%; Pred No. 41;  
 Matches 166; Conservative 0; Mismatches 169; Indels 0; Gaps 0;  
 QY 91 GCCAGGCGATCCAGAGGACCGGGCCACCGTGGCGCGGTGGAGACCGGACCGG 150  
 DB 963 GCCGCGCGCGGAGATGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1022  
 QY 151 CCGAGACGCGGACCTCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 210  
 DB 1023 GCCAGCGCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1082  
 QY 211 GCCTTCGCGCGCGGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 270  
 DB 1083 GGCTGCGCGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTG 1142  
 QY 271 CTGACGAGGAATCGAGTGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 330  
 DB 1143 GTGGGAGGATGATGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1202  
 QY 331 CGCATCATCCCTTAATCTTCGAGCGCGCGGTGAGCGGACAAAGCGCGCGCG 390  
 DB 1203 GCGGTCTGCCACAGTACAGGACCGCGCGCGCGCGCGCGCGCGCGCGCG 1262  
 QY 391 ACCAAGCGGTATCGACCTGTTCCGCGCGGAGTGTG 425  
 DB 1263 GGCCTGACTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1297

RESULT 5  
 OSJN000062  
 LOCUS  
 DEFINITION  
 Oryza sativa genomic DNA, chromosome 4, BAC clone: OSJNB0108J11,  
 complete sequence.  
 ACCESSION  
 AL606618  
 VERSION  
 1  
 KEYWORDS  
 HTG.  
 SOURCE  
 Oryza sativa (japonica cultivar-group)  
 ORGANISM  
 Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Eriatoidae; Oryzaceae; Oryza.  
 REFERENCE  
 AUTHORS  
 Feng,Q., Zhang,Y., Hao,P., Wang,S., Fu,G., Huang,Y., Li,Y., Zhu,J.,  
 Liu,Y., Hu,X., Jia,P., Zhang,Y., Zhao,Q., Ying,K., Yu,S., Tang,Y.,  
 Wang,Q., Zhang,L., Lu,Y., Mu,J., Lu,Y., Zhang,L.S., Yu,Z., Fan,D.,  
 Liu,X., Lu,T., Li,C., Wu,Y., Sun,T., Lei,H., Li,T., Hu,H., Guan,J.,  
 Wu,M., Zhang,R., Zhou,B., Chen,Z., Chen,L., Jin,Z., Wang,R.,  
 Yin,H., Cai,Z., Ren,S., Lv,G., Gu,W., Zhu,G., Tu,Y., Jia,J.,  
 Zhang,X., Chen,J., Kang,H., Chen,X., Shao,C., Sun,Y., Hu,Q.,  
 Zhang,X., Zhang,W., Wang,L., Ding,C., Sheng,H., Gu,J., Chen,S.,  
 Li,L., Zhu,F., Chen,W., Lan,L., Lai,Y., Cheng,Z., Gu,M., Jiang,J.,  
 Li,J., Hong,G., Xue,Y. and Han,B.  
 Sequence and analysis of rice chromosome 4  
 Nature 420 (6913), 316-320 (2002)  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 AUTHORS  
 Fu,G., Wang,S.Y., Ren,S.X., Lv,G., Lin,W., Gu,W.Q., Zhu,G.F.,  
 Tu,Y.F., Jia,J., Yin,H.F., Zhang,Y., Cai,Z., Chen,J., Kang,H.,  
 Chen,X.Y., Shao,Y., Sun,Y., Hu,Q.P., Zhang,X.L., Zhang,W.,  
 Wang,L.J., Ding,C.W., Sheng,H.H., Gu,J.L., Chen,S.T., Ni,L.,  
 Zhu,F.H., Han,B., Feng,Q., Huang,Y.C., Li,Y., Zhu,J.J., Zhao,Q.,  
 Hu,X., Liu,Y.L., Mu,J., Yu,Z., Chen,L., Fan,D.L., Weng,Q.J.,  
 Zhang,L., Lu,Y.Q., Yu,S.L., Liu,X.H., Lu,T., Zhang,Y.J., Lu,Y.,  
 Li,C., Li,T., Zhang,Y., Hu,H., Jia,P.X., Qian,Y.M., Ying,K.,  
 Zhou,B., Chen,Z.H., Hao,P., Zhang,L., Wu,M., Zhang,R.Q., Guan,J.P.  
 and Hong,G.F.  
 Direct Submission  
 Submitted (08-SEP-2001) Han Bin, National Center for Gene Research,  
 Chinese Academy of sciences, 500# Cao Bao Road, Shanghai 200233,  
 CHINA. E-mail enquiries: bhan@ncgr.ac.cn. Clone requests:  
 bhan@ncgr.ac.cn  
 Oryza sativa japonica (nipponbare) genomic DNA, chromosome 4, BAC  
 clone: OSJNB0108J11.  
 On Dec 2, 2003 this sequence version replaced gi:32488349.  
 Web site: http://www.ncgr.ac.cn  
 ----- Summary Statistics  
 Assembly program: phrap  
 -----  
 This is a complete sequence.  
 Genes were identified by a combination of several methods: Gene  
 prediction programs including Fgenesh (<http://www.softberry.com/>),  
 genSCAN (<http://CCR-081.mit.edu/GENSCAN.html>), GeneMarkES  
 (<http://genemark.biology.gatech.edu/Genemark/>), tRNAscan-SE (Sean  
 Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>), searches of the  
 complete sequence against NCBI non-redundant protein database (nr)  
 (<ftp://ncbi.nlm.nih.gov/blast/db/>) and the EST database at NCGR.  
 Location/Qualifiers  
 1..124629  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="genomic DNA"  
 /variety="Nipponbare"  
 /sub\_species="japonica"  
 /db\_xref="taxon:39947"  
 /chromosome="4"  
 /clone="OSJNB0108J11"  
 /clone\_lib="CUGI-OSJNBb"  
 6920..7838  
 /gene="OSJNB0108J11.1"

OSJN000062  
 124629 bp DNA linear  
 Oryza sativa genomic DNA, chromosome 4, BAC clone: OSJNB0108J11,  
 complete sequence.  
 AL606618  
 GI:38605828  
 HTG.  
 Oryza sativa (japonica cultivar-group)  
 Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Eriatoidae; Oryzaceae; Oryza.  
 Feng,Q., Zhang,Y., Hao,P., Wang,S., Fu,G., Huang,Y., Li,Y., Zhu,J.,  
 Liu,Y., Hu,X., Jia,P., Zhang,Y., Zhao,Q., Ying,K., Yu,S., Tang,Y.,  
 Wang,Q., Zhang,L., Lu,Y., Mu,J., Lu,Y., Zhang,L.S., Yu,Z., Fan,D.,  
 Liu,X., Lu,T., Li,C., Wu,Y., Sun,T., Lei,H., Li,T., Hu,H., Guan,J.,  
 Wu,M., Zhang,R., Zhou,B., Chen,Z., Chen,L., Jin,Z., Wang,R.,  
 Yin,H., Cai,Z., Ren,S., Lv,G., Gu,W., Zhu,G., Tu,Y., Jia,J.,  
 Zhang,X., Chen,J., Kang,H., Chen,X., Shao,C., Sun,Y., Hu,Q.,  
 Zhang,X., Zhang,W., Wang,L., Ding,C., Sheng,H., Gu,J., Chen,S.,  
 Li,L., Zhu,F., Chen,W., Lan,L., Lai,Y., Cheng,Z., Gu,M., Jiang,J.,  
 Li,J., Hong,G., Xue,Y. and Han,B.  
 Sequence and analysis of rice chromosome 4  
 Nature 420 (6913), 316-320 (2002)  
 22337377  
 12447439  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 AUTHORS  
 Fu,G., Wang,S.Y., Ren,S.X., Lv,G., Lin,W., Gu,W.Q., Zhu,G.F.,  
 Tu,Y.F., Jia,J., Yin,H.F., Zhang,Y., Cai,Z., Chen,J., Kang,H.,  
 Chen,X.Y., Shao,Y., Sun,Y., Hu,Q.P., Zhang,X.L., Zhang,W.,  
 Wang,L.J., Ding,C.W., Sheng,H.H., Gu,J.L., Chen,S.T., Ni,L.,  
 Zhu,F.H., Han,B., Feng,Q., Huang,Y.C., Li,Y., Zhu,J.J., Zhao,Q.,  
 Hu,X., Liu,Y.L., Mu,J., Yu,Z., Chen,L., Fan,D.L., Weng,Q.J.,  
 Zhang,L., Lu,Y.Q., Yu,S.L., Liu,X.H., Lu,T., Zhang,Y.J., Lu,Y.,  
 Li,C., Li,T., Zhang,Y., Hu,H., Jia,P.X., Qian,Y.M., Ying,K.,  
 Zhou,B., Chen,Z.H., Hao,P., Zhang,L., Wu,M., Zhang,R.Q., Guan,J.P.  
 and Hong,G.F.  
 Direct Submission  
 Submitted (08-SEP-2001) Han Bin, National Center for Gene Research,  
 Chinese Academy of sciences, 500# Cao Bao Road, Shanghai 200233,  
 CHINA. E-mail enquiries: bhan@ncgr.ac.cn. Clone requests:  
 bhan@ncgr.ac.cn  
 Oryza sativa japonica (nipponbare) genomic DNA, chromosome 4, BAC  
 clone: OSJNB0108J11.  
 On Dec 2, 2003 this sequence version replaced gi:32488349.  
 Web site: http://www.ncgr.ac.cn  
 ----- Summary Statistics  
 Assembly program: phrap  
 -----  
 This is a complete sequence.  
 Genes were identified by a combination of several methods: Gene  
 prediction programs including Fgenesh (<http://www.softberry.com/>),  
 genSCAN (<http://CCR-081.mit.edu/GENSCAN.html>), GeneMarkES  
 (<http://genemark.biology.gatech.edu/Genemark/>), tRNAscan-SE (Sean  
 Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>), searches of the  
 complete sequence against NCBI non-redundant protein database (nr)  
 (<ftp://ncbi.nlm.nih.gov/blast/db/>) and the EST database at NCGR.  
 Location/Qualifiers  
 1..124629  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="genomic DNA"  
 /variety="Nipponbare"  
 /sub\_species="japonica"  
 /db\_xref="taxon:39947"  
 /chromosome="4"  
 /clone="OSJNB0108J11"  
 /clone\_lib="CUGI-OSJNBb"  
 6920..7838  
 /gene="OSJNB0108J11.1"





```
C12Q1/68.G01N33/53,
PC GOIN33/566/C12N9/00, (C12P17/10.C12R1.465), C12N15/00, C12N5/00
CC Mitomycin biosynthetic gene cluster
FH Key Location/Qualifiers
FT source 1..53500
FT /organism='Streptomyces lavendulae'.
FEATURES
    source
        Location/Qualifiers
            1..53500
            /organism='Streptomyces lavendulae'
            /mol_type='genomic DNA'
            /db_xref='taxon:1914'
ORIGIN
    Query Match 15.1%; Score 64.4; DB 6; Length 53500;
    Best Local Similarity 51.0%; Pred. No. 17;
    Matches 152; Conservative 0; Mismatches 146; Indels 0; Gaps 0;
QY 92 CCAGCGCGATCCAGAAAGACCGGCGCCACCGCTCGCGCGGTGGGAGACGGCAAGACCGGC 151
DB 43516 CGGCGCGCGGCTCGCGGACCGGTGCGAGATCTTCCGCCGCGACTTCTTCGAGACCATCC 43575
QY 152 CCGACGACGGGACCTCGTTGCCCGCGTGGCCAGTCTCGGCTCGACCTCGACGAG 211
DB 43576 CCGACGCGCGGACGCTTCACTCATCAAGCAGTGTCTGCACGACTGGGACGACGACGACG 43635
QY 212 CCTCGCGCGCGAGTCTGGCGCCCGCGGTCACCCCCGCAGCGACCCCAACCATGGACC 271
DB 43636 TGTACGACATCTCGCGCGGATCGCCACCGCATGAGCCGACTCCCGGCTCTGTGCA 43695
QY 272 TGGACGAGGAATCGAGTGTTCGACCGACCCCAAGCTGGAGACATGAAGCGGC 331
DB 43696 TCGAACAACCTATCGACGAGGCGCGCGCGCATCGACGCTTCTGTCACCTGCTGTC 43755
QY 332 GCATCATCGCCCTAATCTCTGAGCGCCGTGAGCGGACAAAGCGCGCGCATCGAGGA 389
DB 43756 TCGTCTCTGTCGGCGCGCGAAGCTCGGAGAGCGAATTCGCCGCGTCTGTGAGAA 43813
RESULT 9
LOCUS AR266933 53500 bp DNA linear PAT 10-APR-2003
DEFINITION Sequence 76 from patent US 6495348.
ACCESSION AR266933
VERSION AR266933.1 GI:29696315
KEYWORDS
    Unknown.
    Unknown.
ORGANISM
    Unclassified.
REFERENCE
    1 (bases 1 to 53500)
    AUTHORS Sherman,D.H., Mao,Y., Varoglu,M., He,M. and Sheldon,P.
    TITLE Mitomycin biosynthetic gene cluster
    JOURNAL Patent: US 6495348-A 76 17-DEC-2002;
    LOCATION/Qualifiers
    source
        1..53500
        /organism='unknown'
        /mol_type='genomic DNA'
ORIGIN
    Query Match 15.1%; Score 64.4; DB 6; Length 53500;
    Best Local Similarity 51.0%; Pred. No. 17;
    Matches 152; Conservative 0; Mismatches 146; Indels 0; Gaps 0;
QY 92 CCAGCGCGATCCAGAAAGACCGGCGCCACCGCTCGCGCGGTGGGAGACGGCAAGACCGGC 151
DB 43516 CGGCGCGCGGCTCGCGGACCGGTGCGAGATCTTCCGCCGCGACTTCTTCGAGACCATCC 43575
QY 152 CCGACGACGGGACCTCGTTGCCCGCGTGGCCAGTCTCGGCTCGACCTCGACGAG 211
DB 43576 CCGACGCGCGGACGCTTCACTCATCAAGCAGTGTCTGCACGACTGGGACGACGACGACG 43635
QY 212 CCTCGCGCGCGAGTCTGGCGCCCGCGTCACCCCCGCAGCGACCCCAACCATGGACC 271
DB 43636 TGTACGACATCTCGCGCGCGATCGCCACCGCATGAAGCCGACTCCCGGCTCTGTGAGAA 43813
```

```
QY 272 TGGACGAGGAATCGAGTGTTCGCGACCGACCCCAAGCTGGAGACATGAAGCGGC 331
DB 43696 TCGAACAACCTATCGAGCGCGCGCGCATCGAGCTTCTTCGCGACCTCTGCTGC 43755
QY 332 GCATCATCGCCCTAATCTCTGAGCGCGGTGAGCGGACAAAGCGCGCGCATCGAGGA 389
DB 43756 TCGTCTCTGTCGGCGCGCGAAGCTCGGAGAGCGAATTCGCCGCGTCTGTGAGAA 43813
RESULT 10
LOCUS AF127374 63734 bp DNA linear BCT 27-MAY-1999
DEFINITION Streptomyces lavendulae LinA homolog, cytochrome P450 hydroxylase
    CRF3, cytochrome P450 hydroxylase CRF3, MltT (mitT), MltS (mltS),
    MltR (mltR), MltQ (mltQ), MltP (mltP), MltO (mltO), MltN (mltN),
    MltM (mltM), MltL (mltL), MltK (mltK), MltJ (mltJ), MltI (mltI),
    MltH (mltH), MltG (mltG), MltF (mltF), MltE (mltE), MltD (mltD),
    MltC (mltC), MltB (mltB), MltA (mltA), MltCA (mltCA), MltCB (mltCB),
    MltCC (mltCC), MltCD (mltCD), MltCE (mltCE), MltCF (mltCF), MltCG (mltCG),
    MltCH (mltCH), MltCI (mltCI), MltCJ (mltCJ), MltCK (mltCK), MltCL (mltCL),
    MltCM (mltCM), MltCN (mltCN), MltCO (mltCO), MltD (mltD), MltP (mltP), MltQ
    (mltQ), MltR (mltR), MltS (mltS), MltT (mltT), MltU (mltU), MltV
    (mltV), Mlt (mlt), MltW (mltW), MltX (mltX), and MltY (mltY) genes,
    complete cds, and unknown genes.
ACCESSION AF127374.1 GI:4731328
VERSION AF127374
KEYWORDS Streptomyces lavendulae
SOURCE Streptomyces lavendulae
    Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
    Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE
    1 (bases 1 to 63734)
    AUTHORS Mao,Y., Varoglu,M. and Sherman,D.H.
    TITLE Molecular characterization and analysis of the biosynthetic gene
    cluster for the antitumor antibiotic mitomycin C from Streptomyces
    lavendulae NRRL 2564
    JOURNAL Chem. Biol. 6 (4), 251-263 (1999)
    MEDLINE 99201491
    PUBMED 10099135
REFERENCE
    2 (bases 1 to 63734)
    AUTHORS Mao,Y.Q., Varoglu,M. and Sherman,D.H.
    TITLE Direct Submission
    JOURNAL Submitted (10-FEB-1999) Microbiology, Uni. of Minnesota, Box196,
    420 Delaware St. SE., Minneapolis, MN 55455, USA
REFERENCE
    3 (bases 1 to 63734)
    AUTHORS Mao,Y.Q., Varoglu,M. and Sherman,D.H.
    TITLE Direct Submission
    JOURNAL Submitted (27-MAY-1999) Microbiology, Uni. of Minnesota, Box196,
    420 Delaware St. SE., Minneapolis, MN 55455, USA
REMARK Amino acid sequence updated by submitter
FEATURES
    source
        Location/Qualifiers
            1..63734
            /organism='Streptomyces lavendulae'
            /mol_type='genomic DNA'
            /strain='NRRL 2564'
            /db_xref='taxon:1914'
            123..1367
            /function='dehydrogenase'
            /note='ORF6'
            /codon_start=1
            /transl_table=11
            /product='unknown'
            /protein_id='AAD28446.1'
            /db_xref='GI:4731328'
            /translation='MRETLFELGDIPIGTVPQMYASVIRERYGQHPQAPRSEVVD
            VPKVGPGQALVLMAGINYNVWASLGQPDVVISARQKHSEDFHIGSESGVYV
            AVDEGYTOVAVGDEVILSGCQWETADIRLGLADPMTSGSQSVNGYEGYSPAQFAL
            VDGVOCHPKPGLTWEEAACFLLTGATAYRLCGWPHDVRPGDPVLIVGGAGLGSM
            AIOITRAGGIPVAVYDSEERAFYCRFELGQGTINRLDFDHGRLPLDIGHHMGRT
            EGYRAGRFWEVLGHRSPRIVLEHSGGATPTSMYLCNDNAGMVIICGTHGYNADI
            DLRLMWRQRLOGSHFANLRQCRDIHFWANGQLDPCLSWTGGFDIGAKHQLMHDN
            QHPQGNQAVLVNAPRTGLTFFA'
CDS
```



```

Db      50316  CCGGGCGGCGCTCGGCGACCGTGGGAGATCCTCGCGGCGACTTCTTCGAGACCATCC 50375
QY      152    CCGAGCAGCGGACCTCGTGTGCCCGCTGCGCCAGGTGCTCGGCCTCGACCTCGACGAAG 211
Db      50376  CCGAGCGCGCGCGCTACCTCATCAGCACGCTGTCGACACTGGGACGACGACGACG 50435
QY      212    CCTCGCGCGCGAGTCTGCGCCCGCGGTCAACCCCGCCAGCCAGCCACCATGGACC 271
Db      50436  TCGTAGCATCTCCGCGCGATCGCCACCGCATGAACCGGACTCCCGGCTCTGGTCA 50495
QY      272    TGGACGAGAAATCAGAGTGTGTCGACACCGACCCCAAGCTGACGAGGACATGAACGCGC 331
Db      50496  TCGAACACCTCATCGACGAGCGCGCGCGCATCGACGCTCTTCGTCGACCTGCTGCTC 50555
QY      332    GCATCATGCCCTATCTCGGAGCGCGGTGACGCGGACAAAGCGCGCGCGCATCGAGGA 389
Db      50556  TCGTCTCGTGTGCGCGCGCGCAACGCTCGGAGAGCGAATTCGCGCGCTGCTGGAGAA 50613

RESULT 11
LOCUS   AY310323/c 138203 bp DNA linear BCT 06-DEC-2003
DEFINITION Streptomyces sp. FR-008 heptaene macrolide complex synthesis gene
cluster, complete sequence.
ACCESSION AY310323
VERSION   AY310323.1 GI:34766435
KEYWORDS
SOURCE   Streptomyces sp. FR-008
ORGANISM Streptomyces sp. FR-008
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 138203)
AUTHORS  Chen,S., Huang,X., Zhou,X., Bai,L., He,J., Jeong,K.J., Lee,S.Y. and
Deng,Z.
TITLE    Organizational and Mutational Analysis of a Complete
FR-008/Candidin Gene Cluster Encoding a Structurally Related
Polyene Complex
JOURNAL  Chem. Biol. 10 (11), 1065-1076 (2003)
PUBMED  14652074
REFERENCE 2 (bases 1 to 138203)
AUTHORS  Chen,S., Huang,X., Zhou,X., He,J., Bai,L., Jeong,K.J., Lee,S.Y. and
Deng,Z.
TITLE    Direct Submission
JOURNAL  Submitted (30-MAY-2003) Bio-X Life Science Research Center,
Shanghai Jiaotong University, 1954 Huashan Road, Shanghai 200030,
China

FEATURES
source   1..138203
/organism="Streptomyces sp. FR-008"
/mol_type="genomic DNA"
/db_xref="taxon:206662"
complement(574..1950)
/gene="fscO"
complement(574..1950)
/gene="fscO"
/notes="putative FAD-dependent monoxygenase"
/codon_start=1
/transl_table=11
/product="FscO"
/protein_id="AAQ82549.1"
/db_xref="GI:34766436"
/translation="MPSICPAPTSIGRGCPRKGRSRRRIATVTRPCGKVRKH
QAKKRVSMRGSVAVVGGSIGACAAALASRGAEKITVFERADARLDRGVGIGV
HDARVAELRDAGMADENPWFAPLNRRVWSVRDGAHGRAITGTQFFPRAYNWGLWS
ELRRVPSTATYRAGKAVEAQDADGTVRLADGEPPERFDLVIGADGYRVSREAMY
PGTGAAYAGYIGWRGTSFDSGLPSDGLDAHNTFPGGHCWAYRIPDGSGGHRINWVL
YTPARIDGLHDLTPTSLPGRNLNASTLHLRALVAEHFPFWFAAKLLSTPAETTF
IOPIDLVDPVHATDRMWLIGDAASVAREPHLGAGSVKALQATALEAAWAGBSWKEI
LEGYHAAGPVGVTAMVGLARMGSAQVEZTPDWSAMGQAEFDANWQEQNGSDRSRGF
GGHSLKSK"
complement(2264..3037)
/gene="pabC"
complement(2264..3037)

```

```

/gene="pabC"
/notes="putative 4-amino-4-deoxychorismate lyase"
/codon_start=1
/transl_table=11
/product="pabC"
/protein_id="AAQ82550.1"
/db_xref="GI:34766437"
/translation="MIEDGEPAGPEALASLALTNYGHFTTLLVNGRVRGLDLHLER
LIRDTFLDAALDPDRVKRLARRAAPTGRATRVTVDFPALNGLNIADARPGILV
TSRPADKFPGLPRVRSVWRDLPEKSVGLCTPLRLRRQAAGYDDVLYFTGPDGD
ILGEGTNVGLVDGVEVWPGGVLVLAGTTRQLLRGRRHSRHTGRPRGRLSRGLRD
ORRVGVRPVGTIDIREFFPAHSHVTRLAETIYQALPGSL"
complement(3150..3818)
/gene="fscRI"
complement(3150..3818)
/gene="fscRI"
/notes="putative transcriptional activator"
/codon_start=1
/transl_table=11
/product="FscRI"
/protein_id="AAQ82551.1"
/db_xref="GI:34766438"
/translation="MDPAADPAGSPERAANVADGFDAGAVVACLDPSLTITQQNVQ
EFERRFGSSSELCSGRFCDLVHPVQQQLMHQFARMLDGKRHRFATEVIAVQBERTA
STPLNALAVRGRTPDVAAILVVMNAEEAGDADVMAPRKLLSEIDARILEGTA
GVSTIPLASRLYLSRQGVYHVTGLLRKLKVPNRAALVSRAYSMGLVKVTPPKVVE
DFIK"
complement(4377..7205)
/gene="fscRII"
complement(4377..7205)
/gene="fscRII"
/notes="putative transcriptional regulator"
/codon_start=1
/transl_table=11
/product="FscRII"
/protein_id="AAQ82552.1"
/db_xref="GI:34766439"
/translation="MLLERTTEIALVDAALGAATEGRSSVLITGLIGTSALLQRL
SCLADRDVRLRAHAAPMEQDFAGVVRQLFETLLIGSDPEDARWMDAHASFARH
VLADDAAPPQADQAIATBAVLHGLLSLANVSADSKLLIIVDDLOWSDVPSURLTY
LAKRLHGLRAVVCALRDGDPSSHITLAEIREAGTQTLRPASLTATHELVEHFG
EAGDDFVQCHESAENVGNPLFLLSILVGTGFLGRRPLAEHAETARILRSQLERLAS
LIRTPQVRDLAAAIILGEQSDAPTARLAGLSDISYGVALRALGALGALAAPEP
RTHRSVRDAESTLTWQERVHDEAALLYAAGSPAQVAQALMAVTVTRQFVAVD
VLAADATTLRRGPADPTAAGYLCALLDSFPAAGVGRLLVELGTARGFDPCLACERH
IAQAMTLLPPRPDVAALRISPTALGPAPLTAVDLLRQAAEDLGPADGLTGTDRDIA
LRLEARLHGCHEDPHELAASHRLKELGPEPVDVTGGERELTIALGHAATGCHLPA
SEVVRLAGVLDREPATSARIHSTLPLAITMTIADSVEAVDSMLAVEHANEGRTP
ANALVIEQALVHLGRGLAQAREQSETALGTEPAHQGLDTATHTHALVALESDDP
SLAEKVLRRASIRPQGLTTLGILRLLEATREIAGDQGTALDSLSGRLQEAAGWH
NGLVFPWRPWAVGVRRLGDIRTARSLAEFEYARAEQWGAQVGVGRALRLLEDGG
RGIGLLREAVTVLRGSANQLELARLSTLAERLGGGAETEARAATLATSCEPGL
AERARAGAGHRSGPPATPEDVLTPTERRVVVGLACRLTNQEIAGALGVSSRAVEKHLT
HAVRLGISGRRELI VLFSAEYAGTGD"
complement(7210..10254)
/gene="fscRIII"
complement(7210..10254)
/gene="fscRIII"
/notes="putative transcriptional regulator"
/codon_start=1
/transl_table=11
/product="FscRIII"
/protein_id="AAQ82553.1"
/db_xref="GI:34766440"
/translation="MYDLHMPPLELPLERDRRELAALS AVIGELSGRPVAVTVTGEF
GLQGNLDLRLWAAAYATDAGRLVLSHATPAEHEVRYGVVAQLAENRALAPLFLTD
EQPGGLPLNGLLNAVDRPTLLVDRDQWLDPSLRWLEALARLPAPLFTSTT
GTALTPENSVDGPTLGTATIELAPLPTSGGTATVLRFAFGAGDGPAPTEALABAT
RGIPAVVQVLDREPARAGVSPRADRIDPLRALTAEVVVDHATETALSDLRPDAVALR
ALAVCGDLDDFPLVCTILAGHVSSESLRRLAASGLTIRLDGHPRVQDAVVARVLE
EMPAADRAELIAPFAAGLAKRAADDQDLILRLARPVGDPMAVDTLIRGFTSALRG
RDLAVAYLARALDEFLAEDRARIIEFQLASVEMVTAPTAAERRLGGLIRATRPGA
GLRARATDCLLGGDTRAAHRHALAGAI DSAPAPPEPRRGFTGAPVPEPAEPASTPTI

```

```
HLVPHRPVAGTGGATGPDGRRSAAADPLGTPPRAHDAPRSPQPEAAGRPDTAGPDEDP
AVARHELTVLCRVASFURHDDADLDVLPVPALPDVAVNPAVAGVWAEHGVAGVHRE
EVRRLARSAFPAAGTGRLLVVPRLVLLGCRALLTDGDEAENHLGALLDESHREAAI
SVANILTVPAEHLIHBGSPDAARDLAAQAELDLRLHPLPLPVLALALSMTDLONG
HTERARETAAREPLFLAHESATQALLFARGVLARTDDEPHQAREHFRACGRWLRHG
CANPAVQPRVSLAASAAATLGTEBAARLVEELARWRGAAPLGRGAUSLAVTE
ENRNLRAAVTLLSAAARTATRAVILAAAEERNEERTAVAGLPAANTLPPVAP
PGPHVPSAVAPPGPRPAVLGTGRATAKAPGHDAAVWHGLTAVERTTAALAAQGLGNR
EATLAVTTTAVELRLSGVYKLRIRGREELRALVQEAEGS"
complement(10298..13315)
/gene="fscRIV"
/complement(10298..13315)
/gene="fscRIV"
/notes="putative transcriptional regulator"
/codon_start=1
/transl_table=11
/product="FscRIV"
/protein_id="AAQ82554.1"
/db_xref="GI:34766441"
/translation="MWLRWAGDRSAGAAPRGSGSAGPPSLAPSDARESFPVHRRAVP
PAPPLVGRHQDLVLRHADAARAGPRVLVLDGPAGIGKTKTALRAALAEADGPMAGMT
TLVSGRAVDVATGYSGVRLPGLTGRKRTSPLLGGARRALPALAADPGELDA
DPSSTFVLOGLVLYLAANLMDQPLVLDVWDCDSRLWLDPLLRADGLPLLV
AAHRTGTGTAPDALADLVVAHLPLASGLPLDTAEVAELTAHSPPEODPLPSFVGR
LSVTGSGPLEIVRLRLDLSAGLPGDPTGTARLAAGCGKVVAASVGVLPQHPDVRQ
VARAVLGDERTYLAALGRVSVVHVEAEVILRDAGLHHPHLEHSHDIVRAALVD
AVGASGVYLRRAARLLSDVGRSPEEATQLLVPTGDDWAVSLSDAAQAQRG
ACEAAARYLREAREPKDPDVLRLKALAEPTPARVTLTLEHSAITTVRAAAT
AVQLGLTCAVQOQDARLAEALDAELGPEPEPADRELTLESALLIVGSD
KATLPLRLTGLTPQDPTAQQOLAMLSVLSAAGGAEETVSGARRALRAGV
PLGVNSLLTSLALSADENEAEEVLETVLGGSDTAAMTVYVALSTRSLFRLNG
AVPDAAQTLLEILQBRWGDATMTHTAYASVILTERGPDGPRALEALDAIKRPHLD
RFVWEYHWYLMRGAALADGLDGLAQVFGSGASMAEAGLTNEVPLWLEIACILL
SGARGEEAGRAAGTGLAERWGTTRALGYAALARGAAAGTARTGALREAVALLAD
SPARGHAGLGLLGRALVDAGREHREHLEAVGLARRCGVALAARRADELTAAAG
GMREVTAQPLDMLGTGTERTVAGLVASGAGNEVAESLFTVTRVETVELTSLTVYKLVG
ARSGDITLALREAGATAPAAERRPCHAKRRNP"
13522..14898
/gene="fscM"
13522..14898
/gene="fscM"
/notes="glycosyltransferase responsible for mycosamine
attachment"
/codon_start=1
/transl_table=11
/product="FscM"
/protein_id="AAQ82555.1"
/db_xref="GI:34766442"
/translation="WDSARPTLFVSLPESGLNPLLVLAGELARRGVPDLWFATDEH
REEVRLSDVSKVGFASIGVPELSAVTWDDVYREVTSRREKFAHRAVVSYP
ALQARYLEQVDEVDPEALMVDCVAGVGDALARGIPVNLNVPVAGNVLTSHN
PGASVTPKSPFVPSNGLPARMSVRQKLANTLFKWTLGMLPHDMAALLREDAIRK
ELGAPDNMTVRDENAAYVCSVAELDYDPDIDRVSLVGLVLPPLPEAPDDRVTR
WLDQSVYMGFGTITRTEVAAALVEVARMSTGQFQKLPKEQCHLLPAGSL
PNLRVESVPSQDLVLAHPNVSVFFSHGGNAYHEVFGPKVQVRLWVDCFDQAV
RGDFGISTLTKPHVTDPDDVDVKLITRVTDPAFTEAERLGAALQRAAGRAAADI
VTGLLPA"
14953..16011
/gene="fscMII"
14953..16011
/gene="fscMII"
/notes="putative aminotransferase responsible for
mycosamine biosynthesis"
/codon_start=1
/transl_table=11
/product="FscMII"
/protein_id="AAQ82556.1"
/db_xref="GI:34766443"
/translation="WAFTHPVSRPALDGRLELYVSDVSGWISSQGYVRRFEAPA
ENWGAHVACSGTAALTLALRNIGDEVIIVPEFTVWASAWATYTGATPVFVD
CGDNLNIDVTRIEEKTITRAVMPVHVYGRCDMDVMDLALQYNLRVEDSBAHG
VRPVGDIAFCFSLFANKIITVAGEGVCLTDPRLAQLARAFMRDHSFLHKLAY
NYRMTAQAVALAQTERLDEILATREIREARYDAGLKDLPGITLMPARDVLWMDLR
```

```
AERREELRAHLDRAGIETRLFFKPMRSRQGYLDVPMPTLNAHRFESDGLYLPTHGLT
AADQEYITGAVERDFYRAG"
Query Match 15.1%; Score 64.4; DB 1; Length 138203;
Best Local Similarity 49.7%; Pred. No. 13;
Matches 164; Conservative 0; Mismatches 166; Indels 0; Gaps 0;
Qy 13 CCGGGGCTGGGGCCGGCACATCGGCGCATAGTCTCTCACCGCCCGCGAGCGCGCGGA 72
Db 56528 CTGCTGCTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAA 56469
Qy 73 CTGACCAAGAGAGAGAGTTGGCCAGCGCATCCAGAGAGCGCGCGCGCGCGCGCGTGG 132
Db 56468 CTGTCCGAGACCGGTGCGCACGTGACCGTGGAGCGGTGCGACGTCTCCGACCCGCGCG 56409
Qy 133 GAGACCGGCAAGAACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTCT 192
Db 56408 GTCCGGAGAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTCTG 56349
Qy 193 GGCTCGACCTGACGAGAGCCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 252
Db 56348 GAGACCGGCGCGGTGGAATCACTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAA 56289
Qy 253 GCGACCCCAACCATGACCTGAGCGAGGAATCGAGCTGTGCGCGCGCGCGCGCGCGCG 312
Db 56288 GCGGACCGCGCGGTGGAACCTGACGAGGCGCGCGCGCGCGCGCGCGCGCGCGT 56229
Qy 313 GACGAGGACATGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 342
Db 56228 GTCTTCTCTCGATGTCCGCGCATCTCGGC 56199
RESULT 12
CNS08CBV 145539 bp DNA linear PLN 21-NOV-2003
LOCUS Oryza sativa chromosome 12, BAC OSUNBB0075N21 of library OSUNBB
DEFINITION from chromosome 12 of cultivar Nipponbare of ssp. japonica of Oryza
sativa (rice), complete sequence.
ACCESSION AL928758
VERSION AL928758.3 GI:24817647
KEYWORDS HTG.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 145539)
Choisne,N., Orjeda,G., Cattolico,L., Demange,N., Wincker,P.,
Seguins,J.B., Pelletier,E., Scarpelli,C., Salanoubat,M.,
Weissenbach,J. and Quetier,F.
Oryza sativa chromosome 12 sequencing
Unpublished
REFERENCE 2 (bases 1 to 145539)
Genoscope.
Direct Submission
Submitted (20-NOV-2003) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
On Nov 9, 2002 this sequence version replaced gi:23477037.
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: Seqref@genoscope.cns.fr
-----
The following sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : OSJNBa0031E02 (AC-AL954632)
----- Finishing boundaries -----
FINISHED SEGMENT STARTS AT BASE 1
FINISHED SEGMENT ENDS AT BASE 145539
-----
Location/Qualifiers
1. 145539
/organism="Oryza sativa (japonica cultivar-group)"
```

```
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/sub_species="japonica"
/db_xref="taxon:39947"
/chromosome="12"
/clone="OSJNB0075N21"
/clone_lib="OSJNBb"

ORIGIN
Query Match 14.9%; Score 63.4; DB 8; Length 145539;
Best Local Similarity 55.8%; Pred. No. 16;
Matches 121; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 98 GCATCCAGAGACCGGACCGCTCGCGGTGGAGGACGGAAGAACCGGCCGACG 157
Db 34685 GCAGCCAGACTAGCTGTCTCCACAGGCTGGAGCGCTAGTCAAGAGAGCGCCTCG 34744

QY 158 ACGCGGACCTCGTTCGCCCGCTCGCGCCAGGCTCGGCTCGACTCGACGAAAGCCCTCG 217
Db 34745 ACGCGGCGGACGCTGCTCCACTTCGAGCGCTCGGCGGCTCGGCGCCCGCCGCTCT 34804

QY 218 CGCGCGGAGTTCGCGCCCGGCTCACCCCGGACGACCCCAACCATGGAGCTGGACG 277
Db 34805 TCATCGGTGAGGCGCGCGGAGAGCGCGCGCCCGCCCGCCCGCTTCGCGTCA 34864

QY 278 AGGAATCGAGTGTCTCGCGACCGACCCCAAGCTGGA 314
Db 34865 CGCCGACGCGCTCGCTCAACGCGGAGCAGCA 34901

RESULT 13
SC0939115 276800 bp DNA linear BCT 11-FEB-2003
LOCUS Streptomyces coelicolor A3(2) complete genome; segment 12/29.
DEFINITION AL939115 AL035640 AL035654 AL035707 AL049573 AL132674 AL345048
ACCESSION AL355832 AL359989 AL442120 AL596138 AL645882
VERSION AL939115.1 GI:24428505
KEYWORDS Streptomyces coelicolor A3(2)
SOURCE Streptomyces coelicolor A3(2)
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1
AUTHORS Bentley,S.D., Chater,K.F., Cerdeno-Tarraga,A.M., Challis,G.L.,
Thomson,N.R., James,K.D., Harris,D.E., Quail,M.A., Kieser,H.,
Harper,D., Bateman,A., Brown,S., Chandra,G., Chen,C.W., Collins,M.,
Cronin,A., Fraser,A., Goble,A., Hidalgo,J., Hornsby,T., Howarth,S.,
Huang,C.H., Kieser,T., Larke,L., Murphy,L., Oliver,K., O'Neill,S.,
Rabbintwitsch,E., Rajandream,M.A., Rutherford,K., Rutter,S.,
Seeger,K., Saunders,D., Sharp,S., Squares,R., Squares,S.,
Taylor,K., Warren,T., Wietzorrek,A., Woodward,J., Barrall,B.G.,
Parkhill,J. and Hopwood,D.A.
Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2)
JOURNAL Nature 417 (6885), 141-147 (2002)
MEDLINE 21936410
PubMed 12000953
REFERENCE 2 (bases 1 to 276800)
AUTHORS Bentley,S.D.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces
sequencing team, Sanger Institute, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk
COMMENT On or before Oct 30, 2002 this sequence version replaced
GI:4500374, GI:4490616, GI:4490978, GI:6138834, GI:7672242,
GI:7799503, GI:8894754, GI:10241774.
Location/Qualifiers
FEATURES
source
1. 276800
/organism="Streptomyces coelicolor A3(2)"
/mol_type="genomic DNA"
/strain="A3(2)"
/db_xref="taxon:100226"
48..98
stem_loop

/note="Score 59: 21/22 ( 95%) matches, 0 gaps"
complement(146..652)
/gene="SCO3054"
/note="synonym: SCBAC19G2.09c"
complement(146..652)
/gene="SCO3054"
/note="SCBAC19G2.09c, conserved hypothetical protein, len:
168aa: similar to many in Streptomyces coelicolor eg.
TR:Q9F2X0 (EMBL:AL392148) SCD20.09 hypothetical protein
(190 aa) fasta scores; opt: 250, Z-score: 285.6, 34.503%
identity (38.81% ungapped) in 171 aa overlap."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC44519.1"
/db_xref="GI:14970940"
/db_xref="SPTREMBL:Q93J49"
/translation="MSRSPAPGGLGVLTVTLNLTLETGADSLDTEGRAPPLTED
DVFARLRESLRTLLAHAGHPHRAVTPGLGELAAAPLVAVDPAAGSALAPAGE
GSUTARVAARVAQALVAGTWSRLKACEAADCHWAYYDRSPAGRWCSNQVCGARAKM
RRYRAKES"
764..1123
/gene="SCO3055"
/note="synonym: SCBAC19G2.10"
764..1123
/gene="SCO3055"
/note="SCBAC19G2.10, conserved hypothetical protein, len:
119aa: similar to many in Streptomyces coelicolor e.g.
TR:O85701 (EMBL:AF072709) 3SCF60.11c conserved
hypothetical protein (131 aa) fasta scores; opt: 322,
Z-score: 387.9, 42.017% identity in 119 aa overlap and
downstream neighbouring CDS SCBAC19G2.11"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC44520.1"
/db_xref="GI:14970941"
/db_xref="SPTREMBL:Q93J48"
/translation="MPVAELGPVVLDCPPPRALARFYADVLTGTVGEGDWDLRPLPG
GRALYQAAPGFVPVPPAPDDSQPHLDLDKDLDAEKAVLELGARPLDAEDRTGT
FRYVADPAGHPFLCLAC"
786..1114
/gene="SCO3055"
/note="Directly repeated at 11832..12181"
1153..11545
/gene="SCO3056"
/note="synonym: SCBAC19G2.11"
1153..11545
/gene="SCO3056"
/note="SCBAC19G2.11, conserved hypothetical protein, len:
aa: similar to many in Streptomyces coelicolor eg.
TR:Q9F3E9 (EMBL:AL450289) SC2H2.06 hypothetical protein
(150 aa) fasta scores; opt: 340, Z-score: 415.4, 46.032%
identity in 126 aa overlap and upstream neighbouring CDS
SCBAC19G2.10"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC44521.1"
/db_xref="GI:14970942"
/db_xref="SPTREMBL:Q93J47"
/translation="MAPAARFRSVVVDCCDPRELARFYAAGGCTDEADPDWVLQV
PGGPRLSFORAPDLTTPWPWPSDRNACQPHLDFDGGATWAEMDAADHRLVILGARPLD
LEPRDKDFWYADPAGHPFLCLRIEHT"
1178..1527
/gene="SCO3056"
/note="Directly repeated at 11440..11768"
complement(1558..2688)
/gene="SCO3057"
/note="synonym: SCBAC19G2.12c"
complement(1558..2688)
/gene="SCO3057"
/note="SCBAC19G2.12c, possible dipeptidase, len: 376 aa:
repeated_unit

gene
CDS

repeat_unit

gene
CDS

repeat_unit

gene
CDS

repeat_unit

gene
CDS

repeat_unit
```

similar to many eukaryotic peptidases e.g. SW:P31430 (MDPI\_RAT) rat renal dipeptidase (410 aa) fasta scores; opt: 440, Z-score: 474.0, 36.05% identity (38.906% ungapged) in 355 aa overlap. Also weakly similar to TR:Q9KH70 (EMBL:AF268476) thermotable dipeptidase from Brevibacillus borstelensis (307 aa) fasta scores; opt: 273, Z-score: 298.4, 28.159% identity (30.116% ungapged) in 277 aa overlap. Contains pfam match to entry PF01244 Renal dipeptidase. Renal dipeptidase. Also similar to neighbouring upstream CDS SCBAC19G2.13c

/codon\_start=1  
/transl\_table=11  
/product="putative dipeptidase"  
/protein\_id="CAC44522.1"  
/db\_xref="GI:14970943"  
/db\_xref="GOA:Q93J46"  
/db\_xref="STRMBL:Q93J46"  
/translation="MADLQDDLTHTGTEAGGLDLPVLAASAPPEPVAPVSDPDDEP LARAHLAAHPVADGNCGLPWALRHLPWYDELGESAVDTVDPRLRGHVGAALLWSL HLPESLDGDRVAGTLEQLDLVKTIVRAHPGLRAYDAGDAIDARNCRIVAPFGPA GAAALDRIGLIRLGLRLVLTISGVSAEAGLTFGEVRENNRLGVADLSG AGAETVRRTFVAVSKAPALCTRGAAALRHPANLPDILLVELGAAGGLCMVPLTAEO TGPTRVDVADHLHVTAVGQSVGLSGTYDSCAAHPLBLGDPSCYPRIVAEILLRGGW EADVALLTGNVQVLRRAAFTAKAAQLRREPSTATIADLDG"  
/gene="SCO3057"  
/notes="pfam match to entry PF01244 Renal dipeptidase, Renal dipeptidase, score 121.70, E-value 1.4e-32"  
/db\_xref="GI:14970943"  
/db\_xref="GOA:Q93J46"  
/notes="Degenerately, directly repeated at 13490..14610"  
complement(2808..4010)  
/gene="SCO3058"  
/notes="synonym: SCBAC19G2.13c"  
complement(2808..4010)  
/gene="SCO3058"  
/note="SCBAC19G2.13c, possible dipeptidase, len: 400 aa: similar to many eukaryotic peptidases e.g. SW:P31429 (MDPI\_RABIT) rabbit renal dipeptidase (410 aa) fasta scores; opt: 788, Z-score: 879.1, 42.432% identity (48.580% ungapged) in 403 aa overlap. Also weakly similar to TR:Q9KH70 (EMBL:AF268476) thermotable dipeptidase from Brevibacillus borstelensis (307 aa) fasta scores; opt: 450, Z-score: 506.1, 31.034% identity (36.486% ungapged) in 348 aa overlap. Contains pfam match to entry PF01244 Renal dipeptidase. Renal dipeptidase. Also similar to neighbouring downstream CDS SCBAC19G2.12c"  
/codon\_start=1  
/transl\_table=11  
/product="putative dipeptidase"  
/protein\_id="CAC44523.1"  
/db\_xref="GI:14970944"  
/db\_xref="GOA:Q93J45"  
/db\_xref="STRMBL:Q93J45"  
/translation="MTSLKARELLREFFVVDGHNLDLPALREQVRVYDLDDARDIADQ SAHLTDLARLSSGVQAYWYSVRSPLPGAVATATLEQIDCVRLRLDHPGELRAAL TAADMEARAEGRIASLGAEGGHSIDNSLATRALYALGVRYMTLTHNDNNWADSA TDPGVGGLSARVREGRMRENGMLVDLSHVAATMRDALDTAPVIFSHSSSRV CDHPRIIPDVLIERLSANGMAMVTFVPEKVLQAVDVTAEADNNRAHGFHLDSSP EAMKHAPEERVPRTATVSTVADHLHREAVAGVHLGIGGYDGTFFPTPDGLGDV SYPNLILADLRGWSQSLAKLTKWNAVRVLDAAEDVSRGLRAARGSPNATIEQLDGT TAAEQPEG"  
complement(2820..4010)  
/gene="SCO3058"  
/note="pfam match to entry PF01244 Renal dipeptidase, Renal dipeptidase, score 280.40, E-value 2.3e-80"  
complement(2836..3956)  
/note="Degenerately, directly repeated at 12216..13297"  
complement(4025..4567)  
/gene="SCO3059"  
/note="synonyms: pure, SCBAC19G2.14c"  
complement(4025..4567)  
/gene="SCO3059"  
/EC\_number="4.1.1.21"

misc\_feature

repeat\_region

gene

CDS

misc\_feature

repeat\_region

gene

CDS

FEATURES

/note="SCBAC19G2.14c, phosphoribosylaminimidazole carboxylase catalytic subunit PureE, len: 180aa: strongly similar to many eg. SW:Q44679 (PUR6 CORAM) phosphoribosylaminimidazole carboxylase catalytic subunit PureE from Corynebacterium ammoniagenes (177 aa) fasta scores; opt: 666, Z-score: 735.8, 65.541% identity (65.541% ungapged) in 148 aa overlap and TR:O80937 (EMBL:AC004684) putative phosphoribosylaminimidazole carboxylase from Arabidopsis thaliana (645 aa) fasta scores; opt: 644, Z-score: 703.2, 60.870% identity (60.870% ungapged) in 161 aa overlap. Contains Pfam match to entry PF00731 AIRC, AIR carboxylase."

/codon\_start=1

Query Match 14.7%; Score 62.6; DB 1; Length 276800;  
Best Local Similarity 49.5%; Pred. No. 17;  
Matches 161; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

QY 24 GCGCGGCATATGGCCGCATACCTCTACCGCCCGGAGCGCGGACTGACCAAGAG 83  
Db 43436 GGGCGTCGAGAAGCGCGCTGCTGCGCGTCATCGAGCAGATCGGCCCGAGCTGGTCCGCTA 43377

QY 84 CGAGTTGGCCAGCGCGATCCAGAAGGACCGGCGCCACCGTCGCGCGGTGGAGGACCGCAA 143  
Db 43376 CGAGCCACCGAGCGCGCTGATCGACGAGGCATGTTGACCTGGAGCCACCGACAA 43317

QY 144 GAACCGCCCGAGCGCGGACCTCGTTGCCCGTGGCCAGGTGCTGGGCTCGACCT 203  
Db 43316 CAAGGGCTCGCTCGGCGCCAAACGCCATCTCGCGCTCTCCCTCGCGTGGCCACCGCCG 43257

QY 204 CGAGGAGCCCTCGCGCGCGAGTCTGCGCCCGGCTCACCCCGCGAGCGACCCCAAC 263  
Db 43256 CTCGAGCGCGAGCAGCTCGCGCTCTTCGCTACCTGGGCGGCCGAAACCGCACCTGCT 43197

QY 264 CATGGACTCGAGGAGAAATCGAGCTGGTTCGCCACCGCACCCCAAGCTGGAGGAGACAT 323  
Db 43196 GCGGTCGCCATGATGAACATCTGAACGGGCGCTCGACGCGGACTCCAACTGGACAT 43137

QY 324 GAGCGCGCATCATGCCCTATC 348  
Db 43136 CCAGGATTCATGATCGCCCGATC 43112

RESULT 14  
SSBGL3GE SSBGL3GE 2081 bp DNA linear BCT 03-AUG-1994  
LOCUS Streptomyces sp. QM-B814 (AtCC 11238) bg13 gene for beta-glucosidase.  
DEFINITION beta-glucosidase.  
ACCESSION Z29625  
VERSION Z29625.1 GI:525288  
KEYWORDS beta-glucosidase; bg13 gene.  
SOURCE Streptomyces sp.  
ORGANISM Streptomyces sp.  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1 (bases 1 to 2081)  
AUTHORS Perez-Pons J.A., Cayetano, A., Rebordosa, X., Lloberas, J., Guasch, A. and Querol, E.  
TITLE A beta-glucosidase gene (bg13) from Streptomyces sp. strain QM-B814. Molecular cloning, nucleotide sequence, purification and characterization of the encoded enzyme, a new member of family 1 Glycosyl hydrolases  
Eur. J. Biochem. 223 (2), 557-565 (1994)  
JOURNAL MEDLINE 94333346  
PUBMED 8055926  
REFERENCE 2 (bases 1 to 2081)  
AUTHORS Perez-Pons, J.  
TITLE Direct Submission  
JOURNAL Submitted (25-JAN-1994) Josep A. Perez-Pons, Bioquímica i Biologia Molecular, Institut Biologia Fonamental & Dept. Bioquím. i Biol. Mol., Campus Bellaterra. U.A.B., Bellaterra, Barcelona. Catalunya, 08193 Barcelona, Spain  
Location/Qualifiers

```
source
1. 2081
/organism="Streptomyces sp."
/mol_type="genomic DNA"
/strain="QM-B814 (ATCC 11238)"
/db_xref="taxon:1931"
/clone="Clone pJC911"
134..173
repeat_region
134..173
/rpt_type=DIRECT
repeat_unit
134..148
repeat_unit
159..173
RBS
174..180
/standard_names="shine-dalgarno sequence"
/notes="putative"
186..1625
/gene="bgl3"
186..1625
CDS
/gene="bgl3"
/EC_number="3.2.1.21"
/function="hydrolyzes cellobiose and other
beta-D-glucosides"
/standard_names="Streptomyces sp. bgl3 (sspbgl3)"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="beta-glucosidase"
/protein_id="CAA82733.1"
/db_xref="GI:581738"
/db_xref="GOA:O59976"
/db_xref="SPTREMBL:O59976"
/translation="NVPAAQQTATPDAALTFPEGLFNGSATASVOIEGAAEDGRTP
SWDTYTPFVRNGDTGDAIDHHRWEDVALMELGAGVAFSLAWRIQPTGR
GPAQKGDYFYRLADELLAKGIPVATLYHWDLPFLENFGWPFATRFABYAA
IAADALGDKVTKWTTLNEPWCAGFLGSGVGHPTDPVAALRAHNLNHLGLAVQ
ALRDLPLDAQCSVTNLNHHVRPLTDSADADAVRRI DALANRVTFPMQLGAYPEDL
VKDTAGLTDLRVRGDLRLAHOKLDFLGVNYSPTLVSEADSGSTHNSDGHSAHS
PMPGDVAFHPPGPEPTAMGWAVDPGSLVELLRLSDFFALPLVITENGAAPHDYA
DEGNVNDPERIAYVRDHLAAVHRAIKDGSVGRGFLNLSLLDNFENWAGYSKRFQAVY
VDYPTGTRIPKASRWYAEVARTGLPTA"
1676..1719
terminator
/notes="Putative"

ORIGIN
Query Match 14.5%; Score 61.6; DB 1; Length 2081;
Best Local Similarity 50.0%; Pred. No. 89;
Matches 154; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 37 GCGCATACGTCCTCACCGCCGCGGAGCGCGGCGGACTGACCAAGAGCGAGTTGCCAGG 96
Db 651 GCGGAGTACGCGGCATCGCGCGGAGCGCCCTCGCGGACCGCGTGAAGACCTGGACCAC 710
QY 97 CGCATCCAGAGACCGGGCCACCGTCGCGCGGTGGGAGGACGGAGAACCGGCCGAC 156
Db 711 CTCAACGAGCCCTGTGTGACGCGGTTCTCTCGGTACGCGTCCGCGGTGACACGCCCGCGC 770
QY 157 GACGCGGACCTGCTTCCCGCGTGCAGAGTCTCGGCTCGACCTCGACGAAGCCCTC 216
Db 771 CGACACGACCGGTGCGCGCTCGCGCGCGCACCACTCAACTGGGCCACGCGCTG 830
QY 217 GCGCGCGAGGTGTGCGCGCGCGCGGTCAACCGCGCGAGCGACCCCAACCAATGACCTGAC 276
Db 831 GCGGTCCAGCGCTGCGCGACCGCTCCCGCGCGAGCGCGAGTGTCTCGTCAAGCTCAAC 890
QY 277 GAGGAATCGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 336
Db 891 ATCCACACGTACGGCGGTCTACCGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 950
QY 337 ATCGCGCT 344
Db 951 GACGCGCT 958

RESULT 15
BD179992
```

```
LOCUS BD179992 1233 bp DNA linear PAT 15-MAY-2003
DEFINITION Highly thermophilic bacterium-derived protein and gene encoding it.
ACCESSION BD179992
VERSION BD179992.1 GI:30790910
KEYWORDS JP 2002325574-A/483.
SOURCE Thermus thermophilus
ORGANISM Thermus thermophilus
Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
Thermus.
REFERENCE
1 (bases 1 to 1233)
AUTHORS Kuramitsu, N. and Yokoyama, S.
TITLE Highly thermophilic bacterium-derived protein and gene encoding it
JOURNAL Patent: JP 2002325574-A 483 12-NOV-2002;
THE INSTITUTE OF PHYSICAL AND CHEMICAL RESEARCH
COMMENT OS Thermus thermophilus
PN JP 2002325574-A/483
PD 12-NOV-2002
PF 23-FEB-2001 JP 2001116171
PI NARUKI KURAMITSU, SHIGEKUJI YOKOYAMA
PC C12N15/09, C12N15/09, C07K14/195, C12N1/15, C12N1/19, C12N1/21, PC
C12N5/10,
PC C12N9/88, C12P21/02, C12N9/88, C12R1/01, C12N15/09, C12R1/01,
PC C12P21/02, C12R1/01, C12N15/00, C12N15/00, C12N5/00,
PC C12R1/01
CC Highly thermophilic bacterium-derived protein and gene CC
FH Key Location/Qualifiers
FT CDS (1)..(1233).
FEATURES
source
1..1233
/organism="Thermus thermophilus"
/mol_type="genomic DNA"
/db_xref="taxon:274"
ORIGIN
Query Match 14.4%; Score 61.4; DB 6; Length 1233;
Best Local Similarity 47.5%; Pred. No. 1.1e+02;
Matches 182; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY 2 TGCGCAACACACCGGGCTGGGGCGGCGACATGGCGCGCATAGTCTCACCGCCCGG 61
Db 257 TGCGCAAGACGGTGGGGGAGTGGGGTACACCGCGGCAAGTACGGGTTTGACCGCGACA 316
QY 62 AGCGCGCGGACTGACCAAGAGCGAGTTGGCGGCGCATCCAGAGACCGGGCCACCG 121
Db 317 CTGCGCGCTCTCACCGCCATAGACGAGCAGTCCCGCGCATCGCGGGGGTCAACC 376
QY 122 TCGGCGCGGTGGAGGACGGCAAGAACCGGCGCGGACGACGCGGACCTGTTGCCCGGTG 181
Db 377 TCTCTACGAGTGGCGGCTCTCAAGTCCACCGACCCCTGGACCGGGTGGGGGCGGGG 436
QY 182 CCCAGGTCTCGGCTCGACCTCGACGAAGCCCTCGCGCGCGAGTCTGCGCGCCCGCG 241
Db 437 ACCAGGGGCTCATGTTGGTTACGCCACCGACGAGACCCCGAGCTATGCCCTTCCCA 496
QY 242 TCACCGCGCGGACGACCCCAACCATGGACCTGGACGAGGAAATCGAGTGTGCGCACCG 301
Db 497 TCACCGTGGCCACCGGCTCACCATGGCTCGCGGAGTTCGCGAAGACGGGGTTCGCG 556
QY 302 ACCCAAGCTGGACGAGGACATGAGCGCGCGCATATCGCCCTAATCCTGAGCGCGGTG 361
Db 557 CTTACCTCGCGCGCGGCAAGGCCAGGTCACCGTGTCTACAGGGGACAGGCC 616
QY 362 AGCGCGCAAGCGCGCGCGCATC 384
Db 617 TTTACGTCAAGACCGTGGTGGTC 639

Search completed: August 6, 2004, 11:53:10
Job time : 1961.12 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2004, 08:18:49 ; Search time 253.157 Seconds  
(without alignments)  
7148.662 Million cell updates/sec

Title: US-09-855-340A-2  
Perfect score: 426  
Sequence: 1 atgcgaacacacccgggct.....acctttccggcgagctga 426

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002s:\*  
7: Geneseqn2003as:\*  
8: Geneseqn2003bs:\*  
9: Geneseqn2003cs:\*  
10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	426	100.0	426	AAD25932	Aad25932 Micromono
2	426	100.0	4388	AAS08694	Aas08694 Micromono
3	66	15.5	1575	AAQ12342	AAQ12342 Actinomyc
4	64.4	15.1	1056	AAC55832	AAC55832 S. lavend
5	64.4	15.1	1056	ADE10251	Adel0251 S. lavend
6	64.4	15.1	53500	AAC55842	Aac55842 Complete
7	64.4	15.1	53500	ADE10261	Adel0261 S. lavend
8	60.6	14.2	58857	AAA58471	Aaa58471 Nucleotid
9	56.8	13.3	972	ABZ66792	Abz66792 Orthosomy
10	56.8	13.3	45055	ABZ66808	Abz66808 Orthosomy
11	56.2	13.2	2640	ADA69934	Ada69934 Rice gene
12	56.2	13.2	2640	ADC08080	Adc08080 Rice DNA
13	56.2	13.2	113193	AAD54645	Aad54645 Streptomyc
14	56	13.1	662	AAF12486	Aaf12486 Aspergill
15	55.4	13.0	88421	AAL40781	Aal40781 8842int g
16	55	12.9	4770	AAL61172	Aal61172 Actinosyn
17	55	12.9	82746	AAAL6124	Aaal61224 Actinosyn
18	54.2	12.7	1299	ABZ71158	Abz71158 S. muraya
19	54.2	12.7	36321	ABZ71131	Abz71131 Streptomyc
20	53.8	12.6	861	ACA03525	Acc03525 Synthetic
21	53.8	12.6	861	ACC78498	Acc78498 HIV int.o
22	53	12.4	699	ADA70128	Ada70128 Rice gene
23	53	12.4	3072	ACA37796	Aca37796 Prokaryot

24	53	12.4	5811	6	ABS78655	Abs78655 S. macrom
25	52.8	12.4	1266	7	ACA42530	Aca42530 Prokaryot
26	52.8	12.4	8169	5	AAV28609	Aav28609 Actinomad
27	52.8	12.4	109519	5	AAS08693	Aas08693 Micromono
28	52.4	12.3	125401	4	AAD17186	Aad17186 Streptomyc
29	52	12.2	29870	7	AAD36874	Aad36874 Streptomyc
30	51.6	12.1	1185	7	ADA69780	Ada69780 Rice gene
31	51.6	12.1	17083	7	ABZ58812	Abz58812 S. cinnam
32	51.4	12.1	2000	7	ADA71938	Ada71938 Rice gene
33	51.4	12.1	27541	4	AAD17185	Aad17185 Streptomyc
34	51.2	12.0	861	9	ACA03527	Acc03527 Synthetic
35	51.2	12.0	861	9	ADC13260	Adc13260 DNA of HI
36	51.2	12.0	1389	7	ADA71184	Ada71184 Rice gene
37	51	12.0	1527	7	ACA37879	Aca37879 Prokaryot
38	51	12.0	44377	2	AAT80414	Aat80414 Platenoli
39	51	12.0	44377	2	AAT78508	Aat78508 Platenoli
40	50.8	11.9	543	7	ACA37764	Aca37764 Prokaryot
41	50.8	11.9	65140	4	AAD17184	Aad17184 Streptomyc
42	50.6	11.9	1047	7	ABZ66673	Abz66673 Orthosomy
43	50.6	11.9	31263	7	ACA37577	Aca37577 Prokaryot
44	50.4	11.8	1224	3	AAC55788	Aac55788 Hydroxyla
45	50.4	11.8	1224	9	ADE10207	Ade10207 S. lavend

## ALIGNMENTS

## RESULT 1

AAD25932  
ID AAD25932 standard; DNA; 426 BP.

XX  
AC AAD25932;

DT 26-MAR-2002 (first entry)

XX DE Micromonospora carbonacea pMLP1 excisionase DNA.

XX KW Integrase; int; excisionase; xis; integrase attachment site; attP; pMLP1;  
XX KW site-specific integration; hybrid antibiotic; metabolic product;  
XX KW secondary metabolic pathway; ds.

XX OS Micromonospora carbonacea.

XX PN WO200187936-A2.

XX PD 22-NOV-2001.

XX PF 15-MAY-2001; 2001WO-US015760.

XX PR 17-MAY-2000; 2000US-0204670P.

XX (SCHE ) SCHERING CORP.

XX Hosted TJ, Horan AC;

XX WPI; 2002-082983/11.

XX PT Novel polynucleotides encoding integrase, excisionase and an integrase  
XX PT attachment site isolated from a lysogenic phage pMLP1, useful for  
XX PT transforming an actinomycete.

XX PS Claim 5; Page 32-33; 34pp; English.

XX CC The present invention relates to novel polynucleotides encoding integrase  
XX CC (int) and excisionase (xis) and an integrase attachment site (attP) which  
XX CC are isolated from pMLP1, a bacteriophage (lysogenic phage) isolated from  
XX CC Micromonospora carbonacea var. africana. Polynucleotides of the invention  
XX CC are useful for transforming an actinomycete with a vector. They are also  
XX CC useful for creating vectors for site-specific integration into host  
XX CC chromosomes. The integrating vectors are used to express actinomycete  
XX CC genes, manipulate secondary metabolic pathways and create new metabolic  
XX CC products such as hybrid antibiotics. The present sequence is pMLP1  
XX CC excisionase DNA from Micromonospora carbonacea

```
XX SQ Sequence 426 BP; 82 A; 157 C; 143 G; 44 T; 0 U; 0 Other;
Query Match 100.0%; Score 426; DB 6; Length 426;
Best Local Similarity 100.0%; Pred. No. 1.3e-64;
Matches 426; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGCAACACACCGGGCTGGGGCGGCGGCACATGGCGCATACGTCTCTACCGCCGCG 60
Db 1 ATGGCAACACACCGGGCTGGGGCGGCGGCACATGGCGCATACGTCTCTACCGCCGCG 60
Qy 61 GAGCGCCGCGACTGACCAAGAGCGAGTTGGCCAGCGCATCCAGAAAGACCGGGCCACC 120
Db 61 GAGCGCCGCGACTGACCAAGAGCGAGTTGGCCAGCGCATCCAGAAAGACCGGGCCACC 120
Qy 121 GTGGCGCGGTGGAGGAGCGCAAGACCGGACCGCGAGCGGACCTCTGTTCGCCGCTC 180
Db 121 GTGGCGCGGTGGAGGAGCGCAAGACCGGACCGCGAGCGGACCTCTGTTCGCCGCTC 180
Qy 181 GCCCAGGTGCTCGGCTTCGACCTCGACGAAGCCCTCGCGCGCGAGTCTGTGCCCGCGC 240
Db 181 GCCCAGGTGCTCGGCTTCGACCTCGACGAAGCCCTCGCGCGCGAGTCTGTGCCCGCGC 240
Qy 241 GTACCCCGCGAGGACCCCAACCATGGACCTGGAGGAGAAATCGAGTGTTCGCCGAC 300
Db 241 GTACCCCGCGAGGACCCCAACCATGGACCTGGAGGAGAAATCGAGTGTTCGCCGAC 300
Qy 301 GACCCCAAGCTGGACGAGGACATGAAGCGCGCATCATGCCCTTAATCTGGAGCGCGCT 360
Db 301 GACCCCAAGCTGGACGAGGACATGAAGCGCGCATCATGCCCTTAATCTGGAGCGCGCT 360
Qy 361 GAGCGCAAGCGGGCGGCGATCGAGGAAACCAAGCGGCTATCGACTGTTCGCCGCGG 420
Db 361 GAGCGCAAGCGGGCGGCGATCGAGGAAACCAAGCGGCTATCGACTGTTCGCCGCGG 420
Qy 421 AGCTGA 426
Db 421 AGCTGA 426

RESULT 2
ID AAS08694
XX AAS08694 standard; DNA; 4388 BP.
AC AAS08694;
XX
DT 11-SEP-2003 (revised)
DT 26-SEP-2001 (first entry)
XX
XX Micromonospora DNA encoding integrase enzymes.
XX
XX Everninomicin; antibiotic; bottle-neck gene; orthomycin; fermentation;
KW integrase; ds.
XX
XX Micromonospora sp. ATCC 39149.
XX
XX Key Location/Qualifiers
XX CDS 963..1387
XX /tag= a
XX /product= "Integrase #1"
XX 1394..2572
XX /tag= b
XX /product= "Integrase #2"
XX 2570..2799
XX /tag= c
XX /note= "AttB/AttP region of integrase action"
XX 27114..2715
XX /tag= d
XX /label= Insertion_juncture
XX /note= "Site of integrase activity"
XX
XX W0200151639-A2.
```

```
PD 19-JUL-2001.
XX
XX 12-JAN-2001; 2001WO-US001187.
XX
XX 12-JAN-2000; 2000US-0175751P.
XX
XX (SCHE ) SCHERING CORP.
XX
XX Hosted TJ, Horan AC, Wang TX;
XX
XX WPI; 2001-442147/47.
XX
XX P-PSDB; AAU04900, AAU04912.
XX
XX New nucleic acid molecules encoding everninomicin pathway gene products,
XX useful for improving yields of everninomicin, to produce new
XX everninomicin and as probes to identify homologous sequences.
XX
XX Claim 26; Fig 7; 109pp; English.
XX
XX The sequence encodes 2 integrases which permit site specific integration
XX of a vector into an actinomycete, especially a Micromonospora, genome.
XX The invention relates to nucleic acids and vectors comprising a M.
XX carbonacea everninomicin biosynthetic pathway resistance gene product
XX useful for selection of a transfected or transformed host cell. An
XX integrative version of the vector is useful for introducing a
XX everninomicin pathway gene (a bottle-neck gene) into an actinomycete of
XX the genus Micromonospora. The DNA encoding the biosynthetic proteins is
XX useful for synthesizing novel everninomicin-related compounds, arising
XX from modifications of the DNA sequence designed to change glycosyl and
XX modified orsellinic acid groups contained in everninomicin, for
XX expressing functional or mutant everninomicin biosynthetic enzyme for
XX evaluation, diagnosis and preferably biosynthesis of everninomicin or
XX other secondary metabolic products, improving the yield of everninomicins
XX and to produce novel everninomicins and also as a hybridisation probe to
XX identify homologous sequences. The encoded polypeptides are useful for
XX combinatorial biosynthesis to generate libraries of orthomycins, e.g.
XX everninomicin analogues/homologues and drug discovery. The DNA encoding
XX the integrase allows for increasing a given gene dosage. The integrative
XX vector can be used to permanently integrate copies of a heterologous gene
XX of choice into chromosomes of different hosts and to integrate genes
XX which increase the yield of known products or to generate novel products
XX such as hybrid antibiotics or other novel secondary metabolites. The
XX vector can also be used to integrate antibiotic resistance genes. In order
XX to carry out bioconversions with compounds to which the strain is
XX normally sensitive and is thus useful in fermentation processes involving
XX e.g. Streptomyces antibioticus. (Updated on 11-SEP-2003 to standardise OS
XX field)
XX
XX Sequence 4388 BP; 714 A; 1468 C; 1506 G; 694 T; 0 U; 6 Other;
Query Match 100.0%; Score 426; DB 5; Length 4388;
Best Local Similarity 100.0%; Pred. No. 1e-64;
Matches 426; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGCAACACACCGGGCTGGGGCGGCGGCACATGGCGCATACGTCTCTACCGCCGCG 60
Db 963 ATGGCAACACACCGGGCTGGGGCGGCGGCACATGGCGCATACGTCTCTACCGCCGCG 1022
Qy 61 GAGCGCCGCGACTGACCAAGAGCGAGTTGGCCAGCGCATCCAGAAAGACCGGGCCACC 120
Db 1023 GAGCGCCGCGACTGACCAAGAGCGAGTTGGCCAGCGCATCCAGAAAGACCGGGCCACC 1082
Qy 121 GTCGGCCGCGTGGAGGAGCGCAAGAACCGGCGCGGACCGGACCTCTGTTCGCCGCTC 180
Db 1083 GTCGGCCGCGTGGAGGAGCGCAAGAACCGGCGCGGACCGGACCTCTGTTCGCCGCTC 1142
Qy 181 GCCCAGGTGCTCGGCTTCGACCTCGAGAGCCCTCGCGCGCGAGTCTGTTCGCCCGCGC 240
Db 1143 GCCCAGGTGCTCGGCTTCGACCTCGAGAGCCCTCGCGCGCGAGTCTGTTCGCCCGCGC 1202
Qy 241 GTCACCCCGCGCAGCGACCGCCCAACCATGGACCTGGACGAGGAAATCGAGTGTTCGCCGACC 300
Db 1203 GTCACCCCGCGCAGCGACCGCCCAACCATGGACCTGGACGAGGAAATCGAGTGTTCGCCGACC 1262
```

QY 301 GACCCCAAGCTGACGAGGACATGAAGCGCGCATCATCGCCCTAATCTCTGAGCGCCGT 360  
 Db 1263 GACCCCAAGCTGACGAGGACATGAAGCGCGCATCATCGCCCTAATCTCTGAGCGCCGT 1322  
 QY 361 GACCGGACAAAGCGCGCGGATCGAGGAAACCAAGCGGCTCATCGACCTGTTCCGCGG 420  
 Db 1323 GACCGGACAAAGCGCGCGGATCGAGGAAACCAAGCGGCTCATCGACCTGTTCCGCGG 1382  
 QY 421 AGCTGA 426  
 Db 1383 AGCTGA 1388

RESULT 3  
 ID AAQ12342 standard; DNA; 1575 BP.  
 AC AAQ12342;  
 XX  
 DT 24-OCT-2003 (revised)  
 DT 25-MAR-2003 (revised)  
 DT 12-SEP-1991 (first entry)  
 XX  
 DE Actinomycete Phospholipase D gene.  
 XX  
 KW lecithin; actinomycetes; phospholipid; ss.  
 XX  
 OS Streptomyces chromofuscus; A 0848.  
 XX  
 FH Key Location/Qualifiers  
 FT sig\_peptide 1..42  
 FT /\*tag= a  
 FT mat\_peptide 43..1572  
 FT /\*tag= b  
 FT /product= "phospholipase D"  
 XX  
 PN EP435725-A.  
 XX  
 PD 03-JUL-1991.  
 XX  
 PF 11-DEC-1990; 90EP-00403529.  
 XX  
 PR 15-DEC-1989; 89JP-00325355.  
 XX  
 PA (TOXN ) TOYO JOZO KK.  
 PA (ASAH ) ASahi KASEI KOGYO KK.  
 XX  
 PI Yoshioka I, Mizoguchi J, Takahara M, Imamura S, Bappu T;  
 PI Horinouchi S;  
 XX  
 DR WPI; 1991-195237/27.  
 DR P-PSDB; AAR12786.  
 XX  
 PT Cloned DNA encoding phospholipase D - having enzymatic activity and  
 PT produced efficiently in pure form.  
 XX  
 PS Claim 2; Fig 3; 32pp; English.  
 XX  
 CC This sequence was isolated from a gene library prepared from the  
 CC actinomycete Streptomyces chromofuscus A 0848 strain. It was present in  
 CC plasmid pcplD1. See also AAQ12341. (Updated on 25-MAR-2003 to correct PA  
 CC field.) (Updated on 24-OCT-2003 to standardise OS field)  
 XX  
 SQ Sequence 1575 BP; 228 A; 627 C; 514 G; 206 T; 0 U; 0 Other;  
 Query Match 15.5%; Score 66; DB 2; Length 1575;  
 Best Local Similarity 48.4%; Pred. No. 0.0069;  
 Matches 183; Conservative 0; Mismatches 195; Indels 0; Gaps 0;

QY 7 AACACACCGGGTGGCGCGGCACATGGCGCGATAGTCTACCGCGCGCGAGCGC 66  
 Db 646 AAGCGCGCGGTCTCGCGATCTGGGACGACACGAGTTCGCCGACAAACGCTGGTCCGCG 705

QY 67 GCCGAGCTGACCAAGAGGAGTTCGCCAGGCGCATCCAGAAAGACCGGCGCACCGTCCGC 126  
 Db 706 GCGCGGTGAACCAACACCGAGGGCGCGAGGGCACCTGGTCCGCCCTCAAGCGCGCGCC 765  
 QY 127 CGGTGGAGGACGCGCAAGAACCGGCCGACGACGCGGACCTCGTTGCCCGGTGCCCCAG 186  
 Db 766 AAGCAGGCGCTACTTCGAGTGGATCGCGTGGCGCCCGCCCATCGCGGACCACTACGG 825  
 QY 187 GTGCTCGGCTCGACCTCGACGAAGCCCTCGCCCGCGAGGTCTGCGCCCGCGGTACCC 246  
 Db 826 CGGCTGCGCTTCGGCAAGCTCGCGACCTCTCCCTGCTGGACCTCGCTCCCTCCGCTCC 885  
 QY 247 CGCGCAGCGACCCCAACCATGACCTGAGACGAGGAATCGAGCTGGTCCGACCGACCCC 306  
 Db 886 CAGCAGGCGCTCCAGCGCCAGCGTTCGGTGAACGACCGGACCGCTACCTCAGCGCGC 945  
 QY 307 AAGCTGGACGAGGACATGAAGCGCGCATCATCGCCCTAATCTCTGAGCGCGGTGAGCG 366  
 Db 946 GCGCAGCTCGACTGGCTCAGCGCGGCGCTGAAGGCTCGGACACAGGTGGCGGTGGTC 1005  
 QY 367 GACAGGCGCGCGGATC 384  
 Db 1006 GGCAACTCCGTGATGATC 1023

RESULT 4  
 AAC55832  
 ID AAC55832 standard; DNA; 1056 BP.  
 XX  
 AC AAC55832;  
 DT 19-JAN-2001 (first entry)  
 XX  
 DE S. lavendulae MmcR encoding DNA sequence.  
 XX  
 KW Mitomycin; biosynthesis; mitosome ring system; antibiotic; anti-cancer;  
 KW anti-inflammatory; immune-enhancer; immunosuppressant; asthma;  
 KW chronic obstructive pulmonary disease; respiratory inflammation;  
 KW fungicide; pesticide; ds.  
 XX  
 OS Streptomyces lavendulae.  
 XX  
 PN WO200053737-A2.  
 XX  
 PD 14-SEP-2000.  
 XX  
 PR 10-MAR-2000; 2000WO-US006394.  
 XX  
 PR 12-MAR-1999; 99US-00266965.  
 XX  
 PA (MINU ) UNIV MINNESOTA.  
 PA (SHER/) SHERMAN D H.  
 PA (MAOY/) MAO Y.  
 PA (VARO/) VAROGLU M.  
 PA (HENM/) HE M.  
 PA (SHEL/) SHELDOON P C.  
 XX  
 PI Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;  
 XX  
 DR WPI; 2000-601980/57.  
 DR P-PSDB; AAB32531.  
 XX  
 PT Novel nucleic acid molecule comprising mitomycin biosynthetic gene  
 PT cluster useful for cloning mitomycin biosynthetic genes for elucidating  
 PT the molecular basis of mitosome ring system biosynthesis.  
 XX  
 PS Disclosure; Page 281; 399pp; English.  
 XX  
 CC This invention relates to isolated and purified nucleic acid molecules  
 CC from the mitomycin biosynthetic gene cluster. Mitomycins are a group of  
 CC natural products that contain a variety of functional groups, including  
 CC amino benzoquinone and axiridine ring systems. The S. lavendulae

CC Mitomycin biosynthetic gene cluster comprises 47 mitomycin genes spanning  
 CC 55kb of DNA. The invention includes an expression cassette comprising a  
 CC mitomycin biosynthetic gene operably linked to a promoter, and host cells  
 CC transformed with the cassette. The nucleotide, and protein sequences and  
 CC the transformed host cells of the invention result in antiasthmatic,  
 CC anti-inflammatory, cytostatic, immunomodulatory, and antibiotic  
 CC activities. The nucleotide sequences are used to elucidate the molecular  
 CC basis for the biosynthesis of the mitomycin gene system, as well as to  
 CC engineer the biosynthesis of novel natural products, e.g. antibiotics,  
 CC anti-inflammatory agents, anti-cancer agents, immune-enhancers,  
 CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary  
 CC disease as well as other disease involving respiratory inflammation, or  
 CC cholesterol-lowering agents or as crop protection agents (e.g. fungicides  
 CC or insecticides) as well as biopolymers, e.g., in packaging or biomedical  
 CC applications, or to engineer PHA monomer synthases. Sequences AAC55782-  
 CC C55881, AAC55815-C55849 and AAB32485-B32542 represent mitomycin  
 CC biosynthetic gene cluster DNA sequences and encoded proteins. Sequences  
 CC AAC55812-C55814, AAC55850-C55856 and AAC55862-C55869 represent PCR  
 CC primers used in the cloning of the mitomycin biosynthetic genes  
 XX  
 SQ Sequence 1056 BP; 152 A; 427 C; 338 G; 139 T; 0 U; 0 Other;

Query Match 15.1%; Score 64.4; DB 3; Length 1056;  
 Best Local Similarity 51.0%; Pred. No. 0.014;  
 Matches 152; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 92 CCAGGCGCATCCAGAAGACCGCGCCACCGTCCGCGCGTGGGAGGACGCAAGAACCGGC 151  
 DB 683 CCGCGCGCGGCTTCGGGACCGGTCGAGATCTCGCGCGGACTTCTTCGAGACCATCC 742

QY 152 CCGACGACGCGGACCTGTTGCCGCGTCCGCCAGGTCTCGGCTCGACCTCGAGGAG 211  
 DB 743 CCGACGCGCGCGAGCTTACTCATCAAGCACGTCTGCACGACTGGGACGACGACG 802

QY 212 CCTCTCGCGCGCGGAGTCTCGCGCGGCTACCCCGCGAGCGACCCCAACCATGGACC 271  
 DB 803 TCGTACGATCTCCGCGGATCGCCACCGCATGAAGCGGACTCCGCGCTCTGTGTA 862

QY 272 TGGACGAGGAATCGAGCTGTCCGACCGACCCCAAGCTGGACGAGGACATGAACGGC 331  
 DB 863 TCGACACCTCATCGACGAGCGCGCGCGCATCGACGCTCTTCGTCGACCTGCTGCTC 922

QY 332 GCATCATCGCCCTAATCTCTGAGCGCGTGTGAGCGCGCAAGCGCGCGGATCGAGGA 389  
 DB 923 TCGTCTCTGTCGGCGCGCGAAGCGCTCGGAGAGCGAATTCGCGCGCTCTGTGAGAA 980

RESULT 5  
 ADE10251  
 XZ ADE10251 standard; DNA; 1056 BP.

XX AC ADE10251;

XX DT 29-JAN-2004 (first entry)

XX DE S. lavendulae mitomycin biosynthetic protein MmcR gene.

XX KW Mitomycin biosynthetic protein; mitomycin C; antibiotic; MC; apoptosis;  
 XX tumour hypoxia; cytostatic; anti-tumour agent; cancer; gene.

XX OS Streptomyces lavendulae.

XX PN US2003134398-A1.

XX PD 17-JUL-2003.

XX PF 12-SEP-2001; 2001US-00953348.

XX PR 12-SEP-2001; 2001US-00953348.

XX XX (SHER/) SHERMAN D H.

XX PA (MAOY/) MAO Y.

XX PA (VARO/) VAROGLU M.

PA (HEMV/) HE M.  
 PA (SHEL/) SHELTON P.

XX Sherman DH, Mao Y, Varoglu M, He M, Sheldon P;

XX WPI; 2003-863498/80.

XX P-PSDB; ADE10303.

XX New nucleic acid molecule comprising a sequence having mitomycin  
 PT biosynthetic gene cluster, useful for enhancing production of  
 PT antibiotics.

XX Disclosure; SEQ ID NO 66; 308pp; English.

XX The invention relates to an isolated and purified nucleic acid molecule  
 CC comprising a sequence having mitomycin biosynthetic gene cluster, or its  
 CC variant or fragment. Also included are an expression cassette comprising  
 CC the novel nucleic acid molecule (operably linked to a promoter functional  
 CC in a host cell), a recombinant bacterial host cell in which at least a  
 CC portion of a nucleic acid molecule comprising mitomycin biosynthetic gene  
 CC cluster is disrupted (resulting in a recombinant host cell that produces  
 CC altered levels of mitomycin relative to a corresponding nonrecombinant  
 CC bacterial host cell), introducing exogenous DNA into a refractory  
 CC Streptomyces strain, identifying a nucleic acid molecule that is related  
 CC to at least a portion of a nucleic acid molecule comprising a mitomycin  
 CC gene cluster, preparing a compound or its salt from the recombinant host  
 CC cell and a product produced by the recombinant host cell. The nucleic  
 CC acid encodes a MitR, MitS, MitQ, MitP, MitO, MitN, MitM, MitL, MitK,  
 CC MitJ, MitI, MitH, MitG, MitF, MitE, MitD, MitC, MitB, MitA and/or  
 CC MmcA-MmcY. The nucleic acid is useful for enhancing production of  
 CC mitomycin antibiotics, which induce apoptosis and hence are useful as  
 CC anti-tumour (via tumour hypoxia) agents and are useful in treating  
 CC cancer. The gene cluster was isolated from Streptomyces lavendulae. The  
 CC present sequence is a gene encoding a mitomycin biosynthetic protein of  
 CC the invention.

XX Sequence 1056 BP; 152 A; 427 C; 338 G; 139 T; 0 U; 0 Other;

Query Match 15.1%; Score 64.4; DB 9; Length 1056;  
 Best Local Similarity 51.0%; Pred. No. 0.014;  
 Matches 152; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 92 CCAGGCGCATCCAGAAGACCGCGCCACCGTCCGCGCGTGGGAGGACGCAAGAACCGGC 151  
 DB 683 CCGCGCGCGGCTTCGGGACCGGTCGAGATCTCGCGCGGACTTCTTCGAGACCATCC 742

QY 152 CCGACGACGCGGACCTGTTGCCGCGTCCGCCAGGTCTCGGCTCGACCTCGAGGAG 211  
 DB 743 CCGACGCGCGCGAGTCTACTCATCAAGCACGTCTGCACGACTGGGACGACGACG 802

QY 212 CCTCTCGCGCGCGGAGTCTCGCGCGGCTACCCCGCGAGCGACCCCAACCATGGACC 271  
 DB 803 TCGTACGATCTCCGCGGATCGCCACCGCATGAAGCGGACTCCGCGCTCTGTGTA 862

QY 272 TGGACGAGGAATCGAGCTGTCCGACCGACCCCAAGCTGGACGAGGACATGAAGCGC 331  
 DB 863 TCGACACCTCATCGACGAGCGCGCGCGCATCGACGCTCTTCGTCGACCTGCTGCTC 922

QY 332 GCATCATCGCCCTAATCTCTGAGCGCGTGTGAGCGCGCAAGCGCGCGGATCGAGGA 389  
 DB 923 TCGTCTCTGTCGGCGCGCGAAGCGCTCGGAGAGCGAATTCGCGCGCTCTGTGAGAA 980

RESULT 6  
 AAC55842  
 ID AAC55842 standard; DNA; 53500 BP.

XX AC AAC55842;

XX DT 19-JAN-2001 (first entry)

XX DE Complete nucleotide sequence of the mitomycin biosynthetic genes.

KW Mitomycin; biosynthesis; mitosome ring system; antibiotic; anti-cancer;  
KW anti-inflammatory; immune-enhancer; immunosuppressant; asthma;  
KW chronic obstructive pulmonary disease; respiratory inflammation;  
KW fungicide; pesticide; ds.  
XX  
OS Streptomyces lavendulae.  
XX  
XX WO200053737-A2.  
XX  
XX 14-SEP-2000.  
XX  
XX 10-MAR-2000; 2000WO-US006394.  
XX  
XX 12-MAR-1999; 99US-00266965.  
XX  
XX (MINU ) UNIV MINNESOTA.  
XX (SHER/) SHERMAN D H.  
XX (MAO/) MAO Y.  
XX (VARO/) VAROGLU M.  
XX (HEMM/) HE M.  
XX (SHEL/) SHELTON P C.  
XX  
XX Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;  
XX WPI; 2000-601980/57.  
XX  
XX Novel nucleic acid molecule comprising mitomycin biosynthetic gene  
XX cluster useful for cloning mitomycin biosynthetic genes for elucidating  
XX the molecular basis of mitosome ring system biosynthesis.  
XX  
XX Disclosure; Fig 26; 39pp; English.  
XX  
XX This invention relates to isolated and purified nucleic acid molecules  
XX from the mitomycin biosynthetic gene cluster. Mitomycins are a group of  
XX natural products that contain a variety of functional groups, including  
XX amino benzquinone and axiridine ring systems. The S. lavendulae  
XX mitomycin biosynthetic gene cluster comprises 47 mitomycin genes spanning a  
XX 55kb of DNA. The invention includes an expression cassette comprising a  
XX mitomycin biosynthetic gene operably linked to a promoter, and host cells  
XX transformed with the cassette. The nucleotide, and protein sequences and  
XX the transformed host cells of the invention result in antiasthmatic,  
XX antiinflammatory, cytostatic, immunomodulatory, and antibiotic  
XX activities. The nucleotide sequences are used to elucidate the molecular  
XX basis for the biosynthesis of the mitosome ring system, as well as to  
XX engineer the biosynthesis of novel natural products, e.g. antibiotics,  
XX anti-inflammatory agents, anti-cancer agents, immune-enhancers,  
XX immunosuppressants, agents to treat asthma, chronic obstructive pulmonary  
XX disease as well as other disease involving respiratory inflammation, or  
XX cholesterol-lowering agents or as crop protection agents (e.g. fungicides  
XX or insecticides) as well as biopolymers, e.g. in packaging or biomedical  
XX applications, or to engineer PHA monomer syntheses. Sequences AAC55782-  
XX C55881, AAC55815-C55849 and AAB32485-B32542 represent mitomycin  
XX biosynthetic gene cluster DNA sequences and encoded proteins. Sequences  
XX AAC55812-C55814, AAC55850-C55856 and AAC55862-C55869 represent PCR  
XX primers used in the cloning of the mitomycin biosynthetic genes  
XX  
XX Sequence 53500 BP; 7481 A; 19740 C; 19126 G; 7153 T; 0 U; 0 Other;  
XX  
XX Query Match 15.1%; Score 64.4; DB 3; Length 53500;  
XX Best Local Similarity 51.0%; Pred. No. 0.0091;  
XX Matches 152; Conservative 0; Mismatches 146; Indels 0; Gaps 0;  
XX  
XX 92 CCAGGGCGCATCCAGAGGACCGGGCCACCGTGGCGGGTGGGAGGACGCGCAAGACCGGC 151  
XX 43516 CCGGCGCGGCTCGCGGACCGGTGGAGATCTGCCCGGCACTTCTTCGAGACCATCC 43575  
XX  
XX 152 CCGACGACGGGACCTCGTGGCCCGGTGCCAGGTGTCGGCTCGACCTCGACGAG 211  
XX  
XX 43576 CCGACGCGCGGACGCTACCTCATCAAGCAGCTGCTGACGACGTGGAGACGACGACG 43635  
XX  
XX 212 CCTTCGCGCGGAGCTCTGCGCCCGGTACCCCGGACGACCGCCCAACCATGGACC 271  
XX 43636 TGTACGCATCTTCCGCCCGGATCCGACCGCATGAAGCCGGCACTCCCGGCTCTTGTTA 43695

QY 272 TGGACGAGGAATCGAGCTGGTCCGACCGACCCAGCTGGAGGACATGAACGGC 331  
DB 43696 TCGACAACTCATGACGAGCGCGCGCATCGAGCTCTTGTGACCTGTGTCG 43755  
QY 332 GCATCATCGCCCTAATCTGTGAGCGCGGTGAGCGCGCAAGCGCGCGCATCGAGGA 389  
DB 43756 TCGTCTGTGTCGGCGCGCGAACGCTCGGAGAGCGAATTCGCGCGCTGCTGGAGAA 43813

## RESULT 7

ADE10261

ID ADE10261 standard; DNA; 53500 BP.

XX

AC ADE10261;

XX

DT 29-JAN-2004 (first entry)

XX

DE S. lavendulae mitomycin biosynthetic genes complete sequence.

XX

KW Mitomycin biosynthetic protein; mitomycin C; antibiotic; MC; apoptosis;

XX

KW tumour hypoxia; cytostatic; anti-tumour agent; cancer; ds.

XX

OS Streptomyces lavendulae.

XX

PN US2003134398-A1.

XX

PD 17-JUL-2003.

XX

PF 12-SEP-2001; 2001US-00953348.

XX

PR 12-SEP-2001; 2001US-00953348.

XX

(SHER/) SHERMAN D H.

XX

(MAO/) MAO Y.

XX

(VARO/) VAROGLU M.

XX

(HEMM/) HE M.

XX

(SHEL/) SHELTON P.

XX

PI Sherman DH, Mao Y, Varoglu M, He M, Sheldon P;

XX

WPI; 2003-863498/80.

XX

PT New nucleic acid molecule comprising a sequence having mitomycin

XX

PT biosynthetic gene cluster, useful for enhancing production of

XX

PT antibiotics.

XX

PS Disclosure; SEQ ID NO 76; 308pp; English.

XX

CC The invention relates to an isolated and purified nucleic acid molecule comprising a sequence having mitomycin biosynthetic gene cluster, or its variant or fragment. Also included are an expression cassette comprising the novel nucleic acid molecule operably linked to a promoter functional in a host cell, a recombinant bacterial host cell in which at least a portion of a nucleic acid molecule comprising mitomycin biosynthetic gene cluster is disrupted (resulting in a recombinant host cell that produces altered levels of mitomycin relative to a corresponding nonrecombinant bacterial host cell), introducing exogenous DNA into a refractory Streptomyces strain, identifying a nucleic acid molecule that is related to at least a portion of a nucleic acid molecule comprising a mitomycin gene cluster, preparing a compound or its salt from the recombinant host cell and a product produced by the recombinant host cell. The nucleic acid encodes a MitT, MitS, MitR, MitQ, MitP, MitD, MitC, MitB, MitA, and/or MitK, MitJ, MitI, MitH, MitG, MitF, MitE, MitD, MitC, MitB, MitA and/or MmcA-MmcY. The nucleic acid is useful for enhancing production of mitomycin antibiotics, which induce apoptosis and hence are useful as anti-tumour (via tumour hypoxia) agents and are useful in treating cancer. The gene cluster was isolated from Streptomyces lavendulae. The present sequence is an operon or gene cluster encoding the mitomycin biosynthetic proteins of the invention.

XX Sequence 53500 BP; 7481 A; 19740 C; 19126 G; 7153 T; 0 U; 0 Other;

SQ

```
Query Match      15.1%; Score 64.4; DB 9; Length 53500;
Best Local Similarity 51.0%; Pred. No. 0.0091;
Matches 152; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

Qy 92 CCAGCGCATCCAGAGACCGCGCCACCGTCGCGCGTGGAGAGACGACGACCGC 151
Db 43516 CCGCGCGCGCTCGCGGACCGGTGGAGATCTGCGCGCGACTTCTTCGAGACCATCC 43575
Qy 152 CCGACGACGCGGACCTGTTGCCCGGTCCGCCAGTGTCCGCTCGACCTCGACGAG 211
Db 43576 CCGACGCGCGGACGCTTACTCTCATCAAGCACGTCTGCACGACTGGACGACGACG 43635
Qy 212 CCCTCGCGCGCGAGGTCTCGCGCGCGGTCTACCCCGCGAGCGACCCCAACCATGACC 271
Db 43636 TCGTACGATCTCTCGCGCGGTACCGACCGCCATGAAGCGGACTCCCGCTCCTGTCA 43695
Qy 272 TGGACGAGGAATCGAGCTGTGCGACCGACCGACCCCAAGCTGGACGAGACATGAACGCG 331
Db 43696 TCGACAACTTCATCGAGCGCGCGCGCGCATCGACGCTTTCGCGACCTGCTCTGC 43755
Qy 332 GCATCATCGCCCTAATCTCGAGCGCGGTGAGCGCGACGAGCGCGCGGATCGAGGA 389
Db 43756 TCGTCTCTGTCGGCGCGCGGACGCTCGAGAGCGAATTCGCGCGCTGCTGGAGAA 43813

RESULT 8
ID AAA58471 standard; DNA; 58857 BP.
AC AAA58471;
DT 31-OCT-2000 (first entry)
XX Nucleotide sequence of the bleomycin (BLM) gene cluster ORFs 8-30.
XX BLM gene cluster; bleomycin gene cluster; polyketide metabolite;
XX bleomycin; bleomycin analogue; holo-carrier protein; thiazolidine;
XX thiazoline; bithiazoline; microbial metabolite; sugar; ss.
XX Streptomyces verticillus.
XX Key Location/Qualifiers
PH 223..564
CDS /*tag= a
FT /transl_except= (pos: 1..3, aa: Met)
FT /note= "ORF 30; encodes AAB07556"
FT 561..2309
FT /*tag= b
FT /transl_except= (pos: 1..3, aa: Met)
FT /note= "ORF 29; encodes AAB07557"
FT 2787..3486
FT /*tag= c
FT /note= "ORF 28; encodes AAB07558"
FT 3527..5593
FT /*tag= d
FT /transl_except= (pos: 1..3, aa: Met)
FT /note= "ORF 27; encodes AAB07559"
FT 5806..12294
FT /*tag= e
FT /note= "ORF 26; encodes AAB07560"
FT 12291..15491
FT /*tag= f
FT /note= "ORF 25; encodes AAB07561"
FT 15488..21013
FT /*tag= g
FT /note= "ORF 24; encodes AAB07562"
FT 21010..24666
FT /*tag= h
FT /transl_except= (pos: 1..3, aa: Met)
FT /note= "ORF 23; encodes AAB07563"
FT 24663..32690
FT /*tag= i
FT /note= "ORF 22; encodes AAB07564"
FT
```

```
FT 32893..34830
FT /*tag= j
FT /note= "ORF 21; encodes AAB07565"
FT 34827..35804
FT /*tag= k
FT /transl_except= (pos: 1..3, aa: Met)
FT /note= "ORF 20; encodes AAB07566"
FT 35818..37302
FT /*tag= l
FT /transl_except= (pos: 1..3, aa: Met)
FT /note= "ORF 19; encodes AAB07567"
FT 37299..39215
FT /*tag= m
FT /transl_except= (pos: 1..3, aa: Met)
FT /note= "ORF 18; encodes AAB07568"
FT 39301..47181
FT /*tag= n
FT /note= "ORF 17; encodes AAB07569"
FT 47178..49985
FT /*tag= o
FT /transl_except= (pos: 1..3, aa: Met)
FT /note= "ORF 16; encodes AAB07570"
FT 49982..51001
FT /*tag= p
FT /transl_except= (pos: 1..3, aa: Met)
FT /note= "ORF 15; encodes AAB07571"
FT 50998..52386
FT /*tag= q
FT /transl_except= (pos: 1..3, aa: Met)
FT /note= "ORF 14; encodes AAB07572"
FT 52383..52946
FT /*tag= r
FT /note= "ORF 13; encodes AAB07573"
FT 53018..54190
FT /*tag= s
FT /note= "ORF 12; encodes AAB07574"
FT 54187..55824
FT /*tag= t
FT /note= "ORF 11; encodes AAB07575"
FT 55821..56093
FT /*tag= u
FT /transl_except= (pos: 1..3, aa: Met)
FT /note= "ORF 10; encodes AAB07576"
FT 56090..57586
FT /*tag= v
FT /transl_except= (pos: 1..3, aa: Met)
FT /note= "ORF 9; encodes AAB07577"
FT 57583..58857
FT /*tag= w
FT /note= "ORF 8; encodes AAB07578"
XX WO2000040704-A1.
XX 13-JUL-2000.
XX 06-JAN-2000; 2000WO-US0000445.
XX 06-JAN-1999; 99US-0115435P.
XX 05-FEB-1999; 99US-0118484P.
XX 05-JAN-2000; 2000US-00477962.
XX (REGC ) UNIV CALIFORNIA.
XX Shen B, Du L, Sanchez C, Chen M, Edwards DJ;
XX WPI; 2000-465974/40.
XX P-PSDB; AAB07556; AAB07557; AAB07558; AAB07559; AAB07560; AAB07561.
XX AAB07562; AAB07563; AAB07564; AAB07565; AAB07566; AAB07567; AAB07568.
XX AAB07569; AAB07570; AAB07571; AAB07572; AAB07573; AAB07574; AAB07575.
XX AAB07576; AAB07577; AAB07578.
XX New bleomycin gene cluster components useful for peptide and/or
XX polyketide metabolites, especially bleomycin, production and for
PT
```

PT chemically modifying biological molecules.  
XX  
XX  
XX Claim 8; Page 97-136; 162pp; English.  
XX  
CC The present sequence represents the BLM (Bleomycin) gene cluster,  
CC containing open reading frames (ORFs) 8-30. The proteins encoded by the  
CC gene cluster are useful for producing peptides and/or polypeptide  
CC metabolites, especially bleomycin or bleomycin analogues. They are also  
CC useful for chemically modifying biological molecules to produce branched  
CC methyl groups, and for coupling amino acids and fatty acids. They may be  
CC reacted with an apo-carrier protein and coenzyme A to produce a holo-  
CC carrier protein. The BLM gene cluster or catalytic domains can be used  
CC individually or collectively to produce thiazolidine, thiazoline,  
CC bithiazoline and bithiazoline-containing microbial metabolites. The BLM  
CC gene cluster may also be used to produce sugars  
XX  
SQ Sequence 58857 BP; 7256 A; 25139 C; 19353 G; 7109 T; 0 U; 0 Other;  
Query Match 14.2%; Score 60.6; DB 3; Length 58857;  
Best Local Similarity 49.5%; Pred. No. 0.041;  
Matches 156; Conservative 0; Mismatches 159; Indels 0; Gaps 0;  
QY 91 GCCAGCGCATCCAGAAGACCGGGCCACCGCTCGCGCGGTGGAGGCGCAAGAACCGG 150  
DB 8755 GCGCTGCCAAGCGCACGCCCGCGCGGACGCGCGGTACTGTCGCGCGCGCAC 8814  
QY 151 CCGGACGCGGAGACTCTGTTCCCGCGCTGCGCCGAGGTGCTCGGCTCGACTCGACGAA 210  
DB 8815 GGCACCGGAGAGATCTCGCGGCCACCGTCCGGAAGGTGCTGGGCGTGGAGCGGTCCGC 8874  
QY 211 GCCTCTGCGCGCGAGGTCTGCGCGCGCGCTGACCCCGCGAGCACCCCAACATGGAC 270  
DB 8875 ATCGAGCAACTACTTCTGCTGGCGCGGACTCATCCGCGAGCGTATGTCGCCAGC 8934  
QY 271 CTGAGCGAGGAAATCGAGTGTGTCGCGACCGACCCCAAGCTGGACGAGGACATGAAGCGG 330  
DB 8935 CGGGCCCGAGCGCGGGGTGAGGTACCGTGGCGGACCTGCACCGCGCACCCACCGTC 8994  
QY 331 CGCATCATCGCTTAATCTGAGCGCGCTGAGCGCGACGACGCGCGCGCGATCGAGAA 390  
DB 8995 CGGGCTGCGCGCGACCTGACCGCGCGGAGGACCTGCCCGGAGCGCGCGCTCACCGAA 9054  
QY 391 ACCAAGCGGCTCATC 405  
DB 9055 CCTTCGCGCTGATC 9069  
RESULT 9  
ABZ66792  
ID ABZ66792 standard; DNA; 972 BP.  
XX  
XX AC ABZ66792;  
XX  
XX DT 21-MAR-2003 (first entry)  
XX  
DE Orthosomycin biosynthetic polynucleotide SEQ ID NO 246.  
XX  
XX KW Orthosomycin; biosynthesis; everninomicin; avilamycin; gene; ds.  
XX  
XX OS Streptomyces mobaraensis.  
XX  
XX PN WO200279505-A2.  
XX  
XX PD 10-OCT-2002.  
XX  
XX PF 28-MAR-2002; 2002WO-CA000432.  
XX  
XX PR 28-MAR-2001; 2001US-0279095P.  
XX  
XX PR 30-MAR-2001; 2001US-0279709P.  
XX  
XX PR 20-APR-2001; 2001US-0285214P.  
XX  
XX PA (ECOP-) ECOPIA BIOSCIENCES INC.

PI Farnet CM, Zazopoulos E, Staffa A;  
XX  
XX WPI; 2003-058435/05.  
XX  
XX P-PSDB; ABP99329.  
XX  
PT Identifying orthosomycin biosynthetic gene, gene fragment or gene  
PT cluster, by detecting presence of nucleic acid sequence corresponding to  
PT 17 of flambamycins protein families.  
XX  
XX Claim 29; Page 397; 511pp; English.  
XX  
XX The invention relates to identifying orthosomycin biosynthetic genes and  
XX its fragment/gene cluster (ABZ66670-ABZ666813), comprising detecting the  
XX presence of a nucleic acid sequence coding for a polypeptide (ABP99207-  
XX ABP99362). The method is useful for identifying an orthosomycin  
XX biosynthetic gene, gene fragment or gene cluster, especially an  
XX everninomicin-type or avilamycin-type orthosomycin biosynthetic gene,  
XX gene fragment or gene cluster. The method is useful for detecting the  
XX presence of any organism that contains DNA for the production of  
XX orthosomycins (both everninomicin-type orthosomycins and avilamycin-type  
XX orthosomycins) regardless of the level at which genes for orthosomycin  
XX production are expressed by the organism or the amount of new  
XX orthosomycin natural products, not produced by the organism  
XX  
SQ Sequence 972 BP; 145 A; 356 C; 355 G; 116 T; 0 U; 0 Other;  
Query Match 13.3%; Score 56.8; DB 7; Length 972;  
Best Local Similarity 53.6%; Pred. No. 0.27;  
Matches 118; Conservative 0; Mismatches 102; Indels 0; Gaps 0;  
QY 114 GGCCACCGTCCGCGGTGGGAGGACGGCAAGAACCGCGCCCGGACCGACCTCGTTGC 173  
DB 681 GTCCCTGTCGGCGGACCGGATCGTGTGACCTTGGCCCGCGGACCTCGCGGC 740  
QY 174 CCGCGTCCCGAGGTGCTCGGCTCGACCTCGACGAAAGCCCTCGCCCGCGAGGTCTGCG 233  
DB 741 CGCGCGCGGAGTCCGCGCTCACGGACACCGAGCCGAGGTGGACCGCACACCG 800  
QY 234 CCGCGGTACCCCGCGGACCGGACCCCAACCATGACCTGACGAGGAAATCGAGCTGGT 293  
DB 801 CCGGTGACGCGCGGTGACGGACCGGCTGGGGGACCTCTGGAGATCTGGGGAGCT 860  
QY 294 CCGCACCGACCCCAAGCTGGACGAGGACATGAAGCGCGC 333  
DB 861 GCGGACCGCGGCCACGAGGTGGCGGACGTGCTGCTGCGC 900  
RESULT 10  
ABZ66808  
ID ABZ66808 standard; DNA; 45055 BP.  
XX  
XX AC ABZ66808;  
XX  
XX DT 21-MAR-2003 (first entry)  
XX  
XX DE Orthosomycin biosynthetic gene cluster SEQ ID NO 277.  
XX  
XX KW Orthosomycin; biosynthesis; everninomicin; avilamycin; gene; ds.  
XX  
XX OS Streptomyces mobaraensis.  
XX  
XX PN WO200279505-A2.  
XX  
XX PD 10-OCT-2002.  
XX  
XX PF 28-MAR-2002; 2002WO-CA000432.  
XX  
XX PR 28-MAR-2001; 2001US-0279095P.  
XX  
XX PR 30-MAR-2001; 2001US-0279709P.  
XX  
XX PR 20-APR-2001; 2001US-0285214P.  
XX  
XX PA (ECOP-) ECOPIA BIOSCIENCES INC.





FT CDS 70366..79938  
 FT /tag= p  
 FT /product= "Polyketide synthase multienzyme housing  
 FT extension modules 1 and 2 encoded by S. nodosus by amphB  
 FT gene"  
 FT 79956..112709  
 FT CDS  
 FT /tag= q  
 FT /product= "Polyketide synthase multienzyme housing  
 FT extension modules 3, 4, 5, 6, 7 and 8 encoded by S.  
 FT nodosus by amphC gene"  
 FT  
 PN WO200297082-A2.  
 XX  
 XX 05-DEC-2002.  
 XX  
 XX 27-MAY-2002; 2002WO-IE000071.  
 XX  
 XX 31-MAY-2001; 2001IE-00000527.  
 XX  
 XX (UYDU-) UNIV COLLEGE DUBLIN.  
 XX  
 XX Caffrey JP;  
 XX  
 XX WPI; 2003-201271/19.  
 DR P-PSDB; AAE36116, AAE36117, AAE36118, AAE36119, AAE36120, AAE36121,  
 DR AAE36122, AAE36123, AAE36124, AAE36125, AAE36126, AAE36127, AAE36128,  
 DR AAE36129, AAE36130, AAE36131, AAE36132.  
 XX  
 XX Novel cytochrome P450 enzyme and nucleotides encoding the enzyme, useful  
 PT for preparing amphotericin derivative or analog antibiotic agent with  
 PT altered properties, in biosynthesis of polyketide other than  
 PT amphotericin.  
 XX  
 XX Claim 1; Page 52-114; 276pp; English.  
 XX  
 XX The invention relates to the gene cluster encoding the polypeptides  
 CC responsible for the biosynthesis of the polyene antibiotic amphotericin  
 CC (amph) of Streptomyces nodosus. Polynucleotides of the invention are  
 CC useful for preparing amphotericin derivatives or analogue antibiotic  
 CC agents with altered properties and in the biosynthesis of polyketides  
 CC other than amphotericin. amphDII, amphDII or amphDI mutants are useful  
 CC for producing amphotericin derivatives glycosylated with alternative  
 CC sugars; amphDII or amphDI gene sequences are useful in engineered  
 CC biosynthesis of perosaminyl-amphoteronolide B; amphDII or amphDI and  
 CC amphN gene sequences are useful in the engineered biosynthesis of  
 CC perosaminyl-16-desacetoxy-16-methyl amphoteronolide B; amphDII, amphDI  
 CC and amphDI gene sequences are useful for preparing polypeptides capable  
 CC of addition of mycosamine to a polyketide other than amphoteronolide A or  
 CC B or for preparing polypeptides for in vitro synthesis of GDP-mycosamine.  
 CC The present sequence is S. nodosus amph biosynthetic gene cluster  
 XX  
 XX Sequence 113193 BP; 14248 A; 45141 C; 38354 G; 15450 T; 0 U; 0 Other;  
 SQ  
 Query Match 13.2%; Score 56.2; DB 7; Length 113193;  
 Best Local Similarity 46.0%; Pred. No. 0.22;  
 Matches 190; Conservative 0; Mismatches 223; Indels 0; Gaps 0;  
 QY 5 GCAACACCGGGCTGGGGCGGACATATGGGCGCATACGTCTCTACCGCCCGCGAGC 64  
 DB 27792 GCGGCACCGCGCGCATGGGCGGCGCGCGCGCTGGCTCGCGGAGGCGCGCGCC 27851  
 QY 65 GCGCCGGACTGACCAAGAGCGAGTGGCCAGGCGCATCCAGAGGACCGGGCCACCGTCG 124  
 DB 27852 GCTGTGTGTGACCACTCCCGCGCGCAGTGCACACCGAGGAGTGTGGGCGG 27911  
 QY 125 GCGGTGGGAGGACGACAGAACCGCGCCGACGACGCGGACCTGTGTTCGCGCGTCGCC 184  
 DB 27912 AACTCGCGCGGTTCGGCGCGCGAGGTACCGTGGCGCGCGCGGACCGCGGAGC 27971  
 QY 185 AGGTGCTCGGCTCGACCTCGACGAGCCCTCGCGCGCGCAGTCTGCGCCCGCGCGTCA 244  
 DB 27972 CGGCCCGCGCGCTCTCTGTGACGAGTACCGGCGCTACCGCGGTGTGTACGCGGACGACA 28031

QY 245 CCCGCGCAGCGACCCCAACCATGACCTGAGCGAGGAATCGAGCTGGTCCGACCGACC 304  
 DB 28032 CCCGCGGAGCGCCCGCGCGCCCTCGCCCGGTGACACCTTCGCGGAGCGGCCA 28091  
 QY 305 CCAAGCTGGACGAGGACATGAACGCGGCATATCCCTTAATCTCTGAGCGCGGTGAGC 364  
 DB 28092 CCGCGCGCTCCCTGGACGGGTTCGTCTCTTCGGCTCCGTCGCGGGGTCTGGGGCGTGC 28151  
 QY 365 GCGACAGGCGCGCGATGCGAGGAACCAAGCGGCTCATCGACCTGTTCGCGC 417  
 DB 28152 GCGGCCACACCGAGCGCGCGGAGGCGCTATGTGAGCGCCCTCGCGCGC 28204  
 RESULT 14  
 AAF12486  
 ID AAF12486 standard; cDNA; 662 BP.  
 XX  
 AC AAF12486;  
 XX  
 DT 13-MAR-2001 (first entry)  
 XX  
 DE Aspergillus oryzae EST SEQ ID NO:5009.  
 XX  
 KW Multiple gene expression; filamentous fungal cell; EST;  
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
 KW culture condition; environmental stress; spore morphogenesis;  
 KW metabolic pathway engineering; catabolic pathway engineering; ss.  
 XX  
 OS Aspergillus oryzae.  
 XX  
 PN WO200056762-A2.  
 XX  
 PD 28-SEP-2000.  
 XX  
 PP 22-MAR-2000; 2000WO-US007781.  
 XX  
 PR 22-MAR-1999; 99US-00273623.  
 XX  
 PA (NOVO) NOVO NORDISK BIOTECH INC.  
 PA (NOVO) NOVO NORDISK AS.  
 XX  
 PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;  
 XX WPI; 2000-594572/56.  
 DR  
 XX Monitoring differential expression of genes in filamentous fungal cells  
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a  
 PT substrate of expressed sequence tags.  
 XX  
 PS Claim 88; Page 2105; 3161pp; English.  
 XX  
 CC The present invention describes a method for monitoring differential  
 CC expression of genes in a first filamentous fungal (FF) cell relative to  
 CC expression of the same genes in one or more second filamentous fungal  
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
 CC are used in the methods for monitoring differential expression of genes  
 CC in a first filamentous fungal (FF) cell relative to expression of the  
 CC same genes in one or more second filamentous fungal cells. Monitoring the  
 CC global expression of genes from FF cells allows the production potential  
 CC of the microorganisms to be improved. New genes may be discovered,  
 CC possible functions of unknown open reading frames can be identified and  
 CC gene copy number variation and stability can be monitored. The expression  
 CC of genes can be used to study how FF cells adapt to changes in culture  
 CC conditions, environmental stress, spore morphogenesis, recombination,  
 CC metabolic or catabolic pathway engineering. Using ESTs provides several  
 CC advantages over genomic or random cDNA clones including elimination of  
 CC redundancy as one spot on an array equals one gene or open reading frame,  
 CC and organization of the microarrays based on function of the gene  
 CC products to facilitate analysis of the results. AAF07478 to AAF11247  
 CC represents ESTs from Fusarium venenatum; AAF11248 to AAF11853 represents  
 CC ESTs from Aspergillus niger; AAF11854 to AAF14878 represents ESTs from

CC Aspergillus oryzae; and AAF14879 to AAF15337 represents ESTs from  
CC Trichoderma reesei, which are all specifically claimed in the present  
CC invention  
XX  
SQ Sequence 662 BP; 127 A; 257 C; 143 G; 133 T; 0 U; 2 Other;  
Query Match 13.1%; Score 56; DB 3; Length 662;  
Best Local Similarity 51.2%; Pred. No. 0.39;  
Matches 131; Conservative 0; Mismatches 125; Indels 0; Gaps 0;  
QY 45 CGTCTCACCAGCGCGCGGCGGAGTCTGACCAAGAGCGAGTTGGCCAGCGCATCCA 104  
Db 171 CGCCAACTCCCACTTCTACCTCGGCGGAAAGGCGCCACTACTGCCCGGAGAAATCGA 230  
QY 105 GAAGACCGGCGCCAGCGTGGCGCGTGGAGAGCGGCAAGACCGCGCGGAGCGGGA 164  
Db 231 GAAGTGGCGCGCTCCCGCGGCAAGAGACCGCTCTCTCGGCGCAAGTACCTTGA 290  
QY 165 CCGTCTTCCGCGCGTGGCGCGGAGTCTCGGCTCGACCTCGACGAAGCCCTCGCGCGCGC 224  
Db 291 CGTCCCGTCCCGGTGGCGAGCATCTACGTGACCCCGCATGGCGCCCTCTCTTTCAC 350  
QY 225 AGGTCTGGCGCGCGGCGTACCGCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGG 284  
Db 351 CACTCTCTCACTCGGCTACTCTCCCGCGGCTCTCTCGACTGAGGCGTTTCGCGCTAAGCC 410  
QY 285 CGAGCTGGTCCGCACC 300  
Db 411 CGGCAAGACGCGCACC 426  
RESULT 15  
AAL40781  
ID AAL40781 standard; DNA; 88421 BP.  
AC AAL40781;  
XX  
DT 03-OCT-2002 (first entry)  
XX  
DE 88421nt genomic DNA of ramoplanin producing Actinoplanes sp.  
KW Ramoplanin; ramoplanin biosynthetic pathway polypeptide; antibiotic;  
KW biosynthesis gene cluster; biocatalysis; peptide synthetase module;  
KW adenylation domain; hydroxyphenylglycine; HPG; antibiotic precursor;  
KW chlorinate; lipopeptide; gene; ds.  
XX  
OS Actinoplanes sp.  
FH Key Location/Qualifiers  
FT CDS 2077..3078  
FT FT /\*tag= a  
FT FT /product= "Protein of ORF 1"  
FT CDS 3118..4032  
FT FT /\*tag= b  
FT FT /product= "Protein of ORF 2"  
FT CDS 4038..5048  
FT FT /\*tag= c  
FT FT /product= "Protein of ORF 3"  
FT CDS complement(6665..5814)  
FT FT /\*tag= d  
FT FT /product= "Protein of ORF 4"  
FT CDS complement(7703..6693)  
FT FT /\*tag= e  
FT FT /product= "Protein of ORF 5"  
FT CDS complement(9464..8130)  
FT FT /\*tag= f  
FT FT /product= "Protein of ORF 6"  
FT CDS 9691..10761  
FT FT /\*tag= g  
FT FT /product= "Protein of ORF 7"  
FT CDS complement(12751..10829)  
FT FT /\*tag= h  
FT FT /product= "Protein of ORF 8"

FT complement(13617..12802)  
FT FT /\*tag= i  
FT FT /product= "Protein of ORF 9"  
FT CDS complement(15203..13614)  
FT FT /\*tag= j  
FT FT /product= "Protein of ORF 10"  
FT CDS 15591..15863  
FT FT /\*tag= k  
FT FT /product= "Protein of ORF 11"  
FT CDS 15880..19035  
FT FT /\*tag= l  
FT FT /product= "Protein of ORF 12"  
FT CDS 19032..39713  
FT FT /\*tag= m  
FT FT /product= "Protein of ORF 13"  
FT CDS 39713..65800  
FT FT /\*tag= n  
FT FT /product= "Protein of ORF 14"  
FT CDS 65826..66530  
FT FT /\*tag= o  
FT FT /product= "Protein of ORF 15"  
FT CDS 66546..67370  
FT FT /\*tag= p  
FT FT /product= "Protein of ORF 16"  
FT CDS 67384..70059  
FT FT /\*tag= q  
FT FT /product= "Protein of ORF 17"  
FT CDS 70099..70662  
FT FT /\*tag= r  
FT FT /product= "Protein of ORF 18"  
FT CDS 70659..71906  
FT FT /\*tag= s  
FT FT /product= "Protein of ORF 19"  
FT CDS complement(73439..71964)  
FT FT /\*tag= t  
FT FT /product= "Protein of ORF 20"  
FT CDS complement(74216..73563)  
FT FT /\*tag= u  
FT FT /product= "Protein of ORF 21"  
FT CDS complement(75424..74213)  
FT FT /\*tag= v  
FT FT /product= "Protein of ORF 22"  
FT CDS 75535..76464  
FT FT /\*tag= w  
FT FT /product= "Protein of ORF 23"  
FT CDS complement(78110..76449)  
FT FT /\*tag= x  
FT FT /product= "Protein of ORF 24"  
FT CDS complement(79864..78107)  
FT FT /\*tag= y  
FT FT /product= "Protein of ORF 25"  
FT CDS complement(81624..79861)  
FT FT /\*tag= z  
FT FT /product= "Protein of ORF 26"  
FT CDS complement(81909..81682)  
FT FT /\*tag= aa  
FT FT /product= "Protein of ORF 27"  
FT CDS complement(82346..82062)  
FT FT /\*tag= ab  
FT FT /product= "Protein of ORF 28"  
FT CDS 82587..84446  
FT FT /\*tag= ac  
FT FT /product= "Protein of ORF 29"  
FT CDS 84481..85548  
FT FT /\*tag= ad  
FT FT /product= "Protein of ORF 30"  
FT CDS 85556..86845  
FT FT /\*tag= ae  
FT FT /product= "Protein of ORF 31"  
FT CDS 87372..86803  
FT FT /\*tag= af  
FT FT /product= "Protein of ORF 32"  
FT CDS 87494..88420

```

FT FT      /*tag= ag
XX PD      /product= "Protein of ORF 33"
XX PN      WO200231155-A2.
XX XX      18-APR-2002.
XX PD      15-OCT-2001; 2001WO-CA001462.
XX PF      13-OCT-2000; 2000US-0239924P.
XX PR      12-APR-2001; 2001US-0283296P.
XX PR      24-JUL-2001; 2001US-00910813.
XX XX
XX PA      (ECOP-) ECOPIA BIOSCIENCES INC.
XX XX      Farnet CM, Zazopoulos E, Staffa A;
XX XX      WPI; 2002-435445/46.
XX DR      P-PSDB; AAO22146, AAO22147, AAO22148, AAO22149, AAO22150, AAO22151,
XX DR      AAO22152, AAO22153, AAO22154, AAO22155, AAO22156, AAO22157, AAO22158,
XX DR      AAO22159, AAO22160, AAO22161, AAO22162, AAO22163, AAO22164, AAO22165,
XX DR      AAO22166, AAO22167, AAO22168, AAO22169, AAO22170, AAO22171, AAO22172,
XX DR      AAO22173, AAO22174, AAO22175, AAO22176, AAO22177, AAO22178.
XX XX
XX PT      Novel isolated ramoplanin biosynthetic pathway polypeptide useful for
XX PT      chemically modifying biological molecule that is a substrate for a
XX PT      polypeptide encoded by a ramoplanin biosynthesis gene cluster.
XX XX
XX PS      Disclosure; Page 87-135; 212pp; English.
XX XX
XX CC      The invention relates to an isolated ramoplanin biosynthetic pathway
XX CC      polypeptide selected from a polypeptide of open reading frames (ORF) 1-
XX CC      32. The isolated polypeptides are useful for chemically modifying a
XX CC      biological molecule that is a substrate for a polypeptide encoded by a
XX CC      ramoplanin biosynthesis gene cluster, by contacting the biological
XX CC      molecule with the isolated polypeptide, where the polypeptide chemically
XX CC      modifies the biological molecule. The method comprises contacting the
XX CC      biological molecule with at least two different polypeptides encoded by
XX CC      ramoplanin ORFs 1-31. The polypeptides are useful for directing the
XX CC      biosynthesis of the antibiotic ramoplanin in microorganisms. An isolated
XX CC      gene cluster comprising the ORFs is useful as a substrate for
XX CC      bioengineering of antibiotic structures. An isolated polypeptide or its
XX CC      encoding nucleic acid sequence is useful for generating derivatives of
XX CC      ramoplanin for improving production or for producing variants of other
XX CC      antibiotics of the peptide class. The isolated polypeptides are useful
XX CC      for synthesis of ramoplanin in vivo or in vitro, as an adenylation domain
XX CC      in conjunction with other peptide synthetase modules and allowing the
XX CC      incorporation of Thr into a peptide antibiotic precursor, for modifying
XX CC      fatty acid structure and/or enhancing fatty acid incorporation into the
XX CC      peptide antibiotic structure, for production of an hydroxyphenylglycine
XX CC      (HPG)-containing peptide antibiotic, for enhancing secretion of
XX CC      ramoplanin or its variants and derivatives, for enhancing uptake of
XX CC      precursors for ramoplanin biosynthesis, for enhancing production of
XX CC      ramoplanin products or its variants or derivatives, to chlorinate HPG of
XX CC      a peptide antibiotic precursor, and for designing specific nucleotide
XX CC      probes and primers for identifying and isolating putative lipopeptide
XX CC      -producing microorganisms. This polynucleotide sequence represents the
XX CC      88421nt genomic DNA of a ramoplanin producing Actinoplanes sp.
XX CC      microorganism of the invention
XX XX
XX SQ      Sequence 88421 BP; 10639 A; 32391 C; 32663 G; 12728 T; 0 U; 0 Other;
XX XX
XX      Query Match      13.0%; Score 55.4; DB 6; Length 88421;
XX      Best Local Similarity 51.9%; Pred. No. 0.3;
XX      Matches 153; Conservative 0; Mismatches 136; Indels 6; Gaps 1;
XX XX
XX QY      27 CGGCACATGGCGCGCTAGTCTCTACCGCCGCGGAGCGCGCGGACTGACCAAGAGGGA 86
XX DB      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      39811 CGGCACCGTCTGCGCGCGCGCTCGAAACCGCCCTGCGCGGAGCGGCGGCGCA 39870
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY      87 GTTGGCCAGCGCATCCAGAGACCGGGCCACCGTCCGCGCGGAGCGGCGGCAAGAA 146
XX DB      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      39871 GGTGCTGCGTACCGTC-----TATCCGACCGCGCGGAGCGGCGGCGATCCT 39924

```

```

QY      147 CCGGCCCGACGACGCGACCTCGTTCCCGCGCTCGCCAGGTGCTCGGCTCGA 206
DB      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      39925 GCCGCTCGGCGAGACCGGCTTCGGCTCGCGGTGCGGAGGTGACGACGCGGAGTGGA 39984
QY      207 CGAAGCCCTGCGCGCGCGAGGTCTGCGCCCGCGGTACCCCGCCAGCGACCCACCAT 266
DB      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      39985 CGCGCGCGTTCGCGGACGCCACCGCGGACCGCTTCGACCTCGGACCGAGATCCCGGTCCG 40044
QY      267 GGACCTGGACGAGGAAATCGAGCTGGTTCGCGACCGGACCCCAAGCTCGGACGAGGAC 321
DB      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      40045 GGCTCGCTGCTCACCGTTCGAGCGCGGCGGCGGACGTCCTGGCGCTGGTGTGTGCAC 40099

Search completed: August 6, 2004, 09:47:31
Job time : 256.157 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2004, 09:31:14 ; Search time 51.2032 Seconds  
(without alignments)  
4617.079 Million cell updates/sec

Title: US-09-855-340A-2  
Perfect score: 426  
Sequence: 1 atgcgcacacacggggct.....acctgtccgcggagctga 426

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents NA:\*
- 1: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*
  - 2: /cgn2\_6/prodata/2/ina/5B\_COMB.seq:\*
  - 3: /cgn2\_6/prodata/2/ina/6A\_COMB.seq:\*
  - 4: /cgn2\_6/prodata/2/ina/6B\_COMB.seq:\*
  - 5: /cgn2\_6/prodata/2/ina/CTUS\_COMB.seq:\*
  - 6: /cgn2\_6/prodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	64.4	15.1	1056	4	US-09-266-965-66
2	64.4	15.1	5300	4	US-09-266-965-76
3	54.4	12.8	1299	4	US-09-252-991A-5631
4	54.4	12.8	1611	4	US-09-252-991A-5600
5	54.4	12.8	2118	4	US-09-252-991A-5605
6	52	12.2	939	4	US-09-252-991A-3645
7	52	12.2	1881	4	US-09-252-991A-3697
8	52	12.2	1998	4	US-09-252-991A-3812
9	51	12.0	44377	2	US-08-804-227C-7
10	51	12.0	44377	2	US-08-804-198-1
11	50.4	11.8	1224	4	US-09-266-965-22
12	50.4	11.8	12249	4	US-09-266-965-74
13	50.4	11.8	18331	4	US-09-266-965-96
14	50.2	11.8	1344	4	US-09-252-991A-9345
15	50.2	11.8	2112	4	US-09-252-991A-9429
16	50.2	11.8	2454	4	US-09-252-991A-9316
17	50.2	11.8	15872	3	US-09-105-537-1
18	50.2	11.8	15872	4	US-09-091-609-1
19	50.2	11.8	15872	4	US-09-091-609-3
20	50.2	11.8	43280	2	US-08-804-227C-1
21	50.2	11.8	4411529	3	US-09-103-840A-1
22	49.8	11.7	4403765	3	US-09-103-840A-2
23	49.6	11.6	444	4	US-09-252-991A-373
24	49.6	11.6	531	4	US-09-252-991A-426
25	49.6	11.6	7419	4	US-09-252-991A-481
26	49.6	11.6	7449	4	US-09-252-991A-396
27	49.2	11.5	4953	4	US-09-252-991A-5227

C	28	49.2	11.5	5103	4	US-09-252-991A-5192
	29	49.2	11.5	6876	4	US-09-252-991A-5283
	30	49	11.5	1980	4	US-09-252-991A-6896
C	31	48.8	11.5	468	4	US-09-252-991A-3820
	32	48.8	11.5	1281	4	US-09-252-991A-3691
C	33	48.8	11.5	1602	4	US-09-252-991A-3776
	34	48.8	11.5	1755	4	US-09-252-991A-3588
C	35	48.6	11.4	1578	4	US-09-252-991A-5965
	36	48.6	11.4	2469	4	US-09-252-991A-5908
	37	48.6	11.4	3135	4	US-09-252-991A-5922
C	38	48.6	11.4	4411529	3	US-09-103-840A-1
	39	48.4	11.4	765	4	US-09-252-991A-7932
	40	48.4	11.4	1284	4	US-09-252-991A-6826
C	41	48.4	11.4	1305	4	US-09-252-991A-6918
	42	48.4	11.4	1644	4	US-09-252-991A-6420
	43	48.4	11.4	1836	4	US-09-252-991A-7790
	44	48.4	11.4	2088	4	US-09-252-991A-6339
C	45	48.4	11.4	2202	4	US-09-252-991A-6131

ALIGNMENTS

RESULT 1

US-09-266-965-66  
; Sequence 66, Application US/09266965  
; Patent No. 6495348  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D  
; APPLICANT: Mao, Y  
; APPLICANT: Varoglu, M  
; APPLICANT: He, M  
; APPLICANT: Sheldon, P  
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster  
; FILE REFERENCE: 600.456US1  
; CURRENT APPLICATION NUMBER: US/09/266,965  
; CURRENT FILING DATE: 1999-03-12  
; EARLIER APPLICATION NUMBER: US 08/624,447  
; EARLIER FILING DATE: 1996-08-19  
; EARLIER APPLICATION NUMBER: PCT/US94/11279  
; EARLIER FILING DATE: 1994-10-06  
; EARLIER APPLICATION NUMBER: US 08/133,963  
; EARLIER FILING DATE: 1993-10-07  
; NUMBER OF SEQ ID NOS: 145  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 66  
; LENGTH: 1056  
; TYPE: DNA  
; ORGANISM: Streptomyces lavendulae  
US-09-266-965-66

Query Match	15.1%;	Score 64.4;	DB 4;	Length 1056;
Best Local Similarity	51.0%;	Pred. No. 0.00015;		
Matches 152;	Conservative 0;	Mismatches 146;	Indels 0;	Gaps 0;
QY	92	CCAGCGGATCCAGAGGACCGGSCCACCCTGCGCGCGGTGGGAGGAGCGGACCGGC	151	
DB	693	CCGCGCGCGGCTCGCGGACCGGTGCGAGATCTCTCCCGCGGACTTCTTGAGACCATCC	742	
QY	152	CCGACGACGCGGACTCGTTGCGCGCGGTGCGCGGCTCGGCTCGACCTCGACGAAG	211	
DB	743	CCGACGCGCGGACGCTTACTCATCAGACGCTGTCACACTGGGACGACGACGACG	802	
QY	212	CCCTGCGCGCGGAGGTCTGCGCGCGGCTGTCACCGCGGACGACCGACCATGAGACC	271	
DB	803	TCGTACGATCCTCCGCGGATCGCCACCGGCATGAAGCGGACTCCCGGCTCTGTGCA	862	
QY	272	TGGACGAGGAATCGAGTGTTCGCGACCGGACCGACCGAGTGGAGGACATGAGCGGC	331	
DB	863	TCGACACCTCATCGACGCGGCGCGCGGATCGAGCTTCTGTCGACTGCTGTGTC	922	
QY	332	GCATCATCGCCCTAATCTCTGAGCGCGCTGAGCGGCAAGCGCGCGGATCGAGGA	389	

```
Db 923 TCCTCTCTGTCGGCGCGCCGAAACGCTCGAGAGCGAATTCGCCGCGTCTGGAGAA 980

RESULT 2
; Sequence 76, Application US/09266965
; Length: 965-76
; Type: DNA
; ORGANISM: Pseudomonas aeruginosa
; BEST LOCAL SIMILARITY 49.1%; Pred. No. 0.013;
; Mismatches 173; Conservative 0; Gaps 0;
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: Varoglu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600.456US1
; CURRENT APPLICATION NUMBER: US/09/266,965
; CURRENT FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: US 08/624,447
; EARLIER FILING DATE: 1996-08-19
; EARLIER APPLICATION NUMBER: PCI/US94/11279
; EARLIER FILING DATE: 1994-10-06
; EARLIER APPLICATION NUMBER: US 08/133,963
; EARLIER FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 76
; LENGTH: 53500
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
; US-09-266-965-76

Query Match 15.1%; Score 64.4; DB 4; Length 53500;
Best Local Similarity 51.0%; Pred. No. 0.00016;
Matches 152; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 92 CCAGCGGCGATCCAGAGAGCGGCGCCACCGCTCGCGCGTGGGAGGACGCAAGAACCGGC 151
DB 43516 CCGCGCGCGCTCGGCGCGGCGGTCGAGATCTCGCGCGGACTTCTTCGAGACCATCC 43575

QY 152 CCAGAGAGCGGAGCTGTTGCGCGGCTGCGCGGTCGCCAGTGTCTGCGCTCGACCTCGAGGAG 211
DB 43576 CCAGCGCGCGGCGGCTGCTACCTCATCAAGCAGCTGCTGACGACTGGGAGGAGCGAGCG 43635

QY 212 CCCTCGCGCGCGGCGGCTGCTGCGCGCGGCTGCTGCGCGGCGGCGGCGGCGGCGGCGGCGG 271
DB 43636 TCCTAGCGATCTCTCGCGCGGATCGCCACCGCATGAGCGGACTCCCGCTCTCTGTTCA 43695

QY 272 TGACAGAGGAAATCGAGTGTGTCGCGCGGACCGACCGACCGGAGTGGAGGAGCATGAGCGGC 331
DB 43696 TCACAACTCATCGAGAGCGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 43755

QY 332 GCATATCGCCTTAATCTGAGCGCGGCTGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 389
DB 43756 TCCTCTCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 43813

RESULT 3
; Sequence 5631, Application US/09252991A
; Length: 5631-5631
; Type: DNA
; ORGANISM: Pseudomonas aeruginosa
; BEST LOCAL SIMILARITY 49.1%; Pred. No. 0.013;
; Mismatches 173; Conservative 0; Gaps 1;
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5631
; LENGTH: 1611
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-5631

Query Match 12.8%; Score 54.4; DB 4; Length 1611;
Best Local Similarity 49.1%; Pred. No. 0.013;
Matches 173; Conservative 0; Mismatches 176; Indels 3; Gaps 1;

QY 63 GCGCGCGGCGGACTGACCAAGAGCGGAGTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 122
DB 1216 GGGCGTGGTGGGATCATCTACGAGTCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1157
QY 123 GCGCGCGTGGGAGGAGCGGCAAGAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 182
DB 1156 GTGCGTGAAGTCCGGCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1097
QY 183 CCAGTGTCTGGCTTCGACCTCGACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 242
DB 1096 CCAGGCGATCGCGGCTGTCATCCAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1040
QY 243 CACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 302
```

```
; LENGTH: 1299
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-5631

Query Match 12.8%; Score 54.4; DB 4; Length 1299;
Best Local Similarity 49.1%; Pred. No. 0.013;
Matches 173; Conservative 0; Mismatches 176; Indels 3; Gaps 1;

QY 63 GCGCGCGGCGGACTGACCAAGAGCGGAGTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 122
DB 390 GGGCGTGGTGGGATCATCTACGAGTCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 449
QY 123 GCGCGCGTGGGAGGAGCGGCAAGAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 182
DB 450 GTGCGTGAAGTCCGGCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 509
QY 183 CCAGTGTCTGGCTTCGACCTCGACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 242
DB 510 CCAGGCGATCGCGCGCTGTCATCCAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 566
QY 243 CACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 302
DB 567 GCGCGTGCAGTGTGGAAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 626
QY 303 CCGCAAGCTGGACGAGGAGCATGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 362
DB 627 CGAATAGCTGACGTGATCGTCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 686
QY 363 GCGCGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 414
DB 687 CGAGGCGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 738

RESULT 4
; US-09-252-991A-5600/c
; Sequence 5600, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5600
; LENGTH: 1611
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-5600

Query Match 12.8%; Score 54.4; DB 4; Length 1611;
Best Local Similarity 49.1%; Pred. No. 0.013;
Matches 173; Conservative 0; Mismatches 176; Indels 3; Gaps 1;

QY 63 GCGCGCGGCGGACTGACCAAGAGCGGAGTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 122
DB 1216 GGGCGTGGTGGGATCATCTACGAGTCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1157
QY 123 GCGCGCGTGGGAGGAGCGGCAAGAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 182
DB 1156 GTGCGTGAAGTCCGGCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1097
QY 183 CCAGTGTCTGGCTTCGACCTCGACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 242
DB 1096 CCAGGCGATCGCGGCTGTCATCCAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1040
QY 243 CACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 302
```

Db 1039 CGCGCTGAGGTGGTGAACACCGACCGCGCGGTGCGCGCTGATCAGATGCC 980  
QY 303 CCCCAAGCTGACGACGAGGATCAAGACGGCGCATCATCCCTTAATCTTGGAGCCCGTGA 362  
Db 979 CGAATACGTGACGTGATCTCCCGCGCGGCAAGGCGCTGATCGAGCGCATCAGCG 920  
QY 363 GCGGACAAAGCGCGCGCATCGAGGAACCAAGCGGCTCATCGACCTGTC 414  
Db 919 CGAGCCAAAGTGCCTGATCAAGCATCTGGACGGCATCTGCCACGCTTAC 868

## RESULT 5

US-09-252-991A-5605  
; Sequence 5605, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 5605  
; LENGTH: 2118  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-5605

Query Match 12.8%; Score 54.4; DB 4; Length 2118;  
Best Local Similarity 49.1%; Pred. No. 0.013;  
Matches 173; Conservative 0; Mismatches 176; Indels 3; Gaps 1;  
QY 63 GCGCGCGGACTGACCAAGCGAGTGGCCAGGCGCATCCAGAGGACCGGCGCCACCGT 122  
Db 565 GGGGTGTCGGGATCATCTACGATCGCGCGGACGATGACATCGACGCGCCAGCT 624  
QY 123 CGCGCGGTGGAGGACGGCAAGAACCGCGCGACGAGCGGACCTGTTGCCCGGTGCG 182  
Db 625 GTGCTGAAATCGGCAAGCGGACCATCTGCGCGCGGCTCCGAGGCGATCCATCAA 684  
QY 183 CCAGGTCTCGGCTCGACTCGACGAAGCCCTCGCGCGCGAGTCTCGCGCCCGGGT 242  
Db 685 CCAGGCGATCGCCGCTGATCGACGAGGCGCTGCGCGAGCGCGCTG---CCGCGCG 741  
QY 243 CACCCCGCAGCGACCCCAACCATGACTGAGAGGAAATCGAGCTGCTGTCGACCGA 302  
Db 742 CGCGGTGACGTGGTGGAAACACACCGCGCGCGGCTGCGCGGCTGATCAGCATGCC 801  
QY 303 CCCCAAGCTGACGAGGACATGAAGCGGCGCATCTCGCCCTAATCTGAGGCGGTGA 362  
Db 802 CGAATACGTGACGTGATCGTCCCGCGCGGCGGCAAGGCGCTGATCGAGCGCATAGCG 861  
QY 363 GCGGCAAAAGCGCGCGCGCATCGAGGAACCAAGCGGCTCATCGACCTGTC 414  
Db 862 CGAGGCCAAGTCCCGTGTATCAAGCATCTGGACGGCATCTGCCAGCTTAC 913

## RESULT 6

US-09-252-991A-3645  
; Sequence 3645, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 3645  
; LENGTH: 939  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-3645

Query Match 12.2%; Score 52; DB 4; Length 939;  
Best Local Similarity 47.3%; Pred. No. 0.039;  
Matches 194; Conservative 0; Mismatches 210; Indels 6; Gaps 1;

QY 14 CGGGGTGGCGCGGCACATGGCGCGCATACCTCTCACCGCCCGGAGCGCGCGGAC 73  
Db 348 CGTGCCTGGCGGCGCGGCTCTGAGCGCTTACGTGGCGACATCGACCGTTTCTGGCA 407  
QY 74 TGACCAAGAGCGAGTTTGGCCAGCGCATCCAGAAAGCACCGGCGCACCGTCTGGCGGTGG 133  
Db 408 TCCTCCAGGCCAAGGCGGCGGCTCGCAACTGCTCAGCAGGTGCGAGGCGCTGTGCC 467  
QY 134 AGGACGGCAAGAACCGGCGCGGACGCGGACCTCTGTCGCCGCTGCGCGGAGTCTCG 193  
Db 468 TGTTCCTCAGCCTGCTGTGTGTGTCGTACCCCTCTACGACCTCAGTATCAGTCTATCG 527  
QY 194 GCCTCGACCTCGACGAAGCCCTCGCGCGCGAGTCTGCGCCCGGCGGTCAACCGCGCAG 253  
Db 528 GCGCGTGGCGGACTCACCAGCACCGCGCGCGCTGGCGCGGCGGAGCTGGACGCGC 587  
QY 254 CGACCCCAACCATGACCTCGAGGAGAAATCGAGTGTGTCGCGACCGACCCCAAGTGG 313  
Db 588 GGGTCACCTACAGCGCGGAGGACCACTGGCGCAACTGGCGGAGCGGCTTCAACCAATGG 647  
QY 314 ACAGGACATCAGCGCGCATCATCGCCCTTAATCTTGGAGCGCGGTGAGCGGACAAAG 373  
Db 648 CCGCGAGCTGAAGAG-----CATCTACGCGGACCTGGAAGACCGGCTGGAGACAGA 701  
QY 374 CGCGCGGATCGAGGAACCAAGCGGTCTATCGACCTGTTTCCGCGCGAGC 423  
Db 702 CCGCGCGCTCTCGACAGCACCCAGCGCTCGAATGCTCTAGCGCAGC 751

## RESULT 7

US-09-252-991A-3697  
; Sequence 3697, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 3697  
; LENGTH: 1881  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-3697

Query Match 12.2%; Score 52; DB 4; Length 1881;  
Best Local Similarity 47.3%; Pred. No. 0.039;  
Matches 194; Conservative 0; Mismatches 210; Indels 6; Gaps 1;  
QY 14 CGGGCTGGCGCGGCACATGGCGCGCATACCTCTCACCGCGGAGCGCGCGGAC 73



QY 74 TGACCAAGAGCGAGTTGCCAGGCGCATCCAGAAAGACCGGGCCACCGCTCGCGCGGTGGG 133  
Db 26009 TCGTCATCGAGAGAGCGCGCGAGGCCACCGGCCCGAACTGCTCGCGCCGAAACCGCGCG 26068  
QY 134 AGGACGGCAAGAACCGGCCGACGACGCGGACCTCGTTGCCCGGTCCGCCAGGTGCTCG 193  
Db 26069 CCGAGCGGACGCTGCTGGTCCGAGGAGTGTGGCAGAGTGACCGTGGCCCTGATGATGT 26128  
QY 194 GCGTGCACCTCGAAGAGCCCTCGCGCGCGCGAGGTCTGGCGCCCGCGGTCAACCGGCCAG 253  
Db 26129 CCGCGCACAAACGCGCCCTGCGCGACCGAGCGCGCGCGCTGCGCGCGCGACCTGCTCG 26188  
QY 254 CGACCCCAACCACTGACCTGGAACGAGGAATCGAGTGTGTCGACCGACCGCCCAAGCTGG 313  
Db 26189 CCCACCCGAGCTGACCGCGCGCGAGCTCGGCTACACCTCATCACCCGCGACCGCGT 26248  
QY 314 ACGAGGA 320  
Db 26249 TCGAGCA 26255

RESULT 10  
US-08-804-198-1  
; Sequence 1, Application US/08804198  
; Patent No. 5945320  
; GENERAL INFORMATION:  
; APPLICANT: Burgett, Stanley G.  
; APPLICANT: Kuhstoss, Stuart A.  
; APPLICANT: Rao, Nagaraja R.  
; APPLICANT: Richardson, Mark A.  
; APPLICANT: Rostek, Paul R., Jr.  
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PAUL R. CANTRELL, 1138  
; STREET: LILLY CORPORATE CENTER  
; CITY: INDIANAPOLIS  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Macintosh  
; OPERATING SYSTEM: Macintosh 7.0  
; SOFTWARE: Microsoft Word 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/804,198  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CANTRELL, PAUL R.  
; REGISTRATION NUMBER: 36,470  
; REFERENCE/DOCKET NUMBER: P9113  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-3885  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 44377 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 350...14002  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 14046...20036  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 20110...31284  
; FEATURE:

; NAME/KEY: CDS  
; LOCATION: 31329...36071  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 36155...41830  
; US-08-804-198-1  
Query Match 12.0%; Score 51; DB 2; Length 44377;  
Best Local Similarity 47.9%; Pred. No. 0.065;  
Matches 147; Conservative 0; Mismatches 160; Indels 0; Gaps 0;  
QY 14 CGGGGCTGGGGCGCGGCACATGCGCGCATAGCTCTCAGCGCCGCGAGCGCGCGGAC 73  
Db 25949 CCGAGGGGCGCGCGCGCGCGGCATCTCTCTGTTGCGCATCAGCGGCACCAACGCGCACC 26008  
QY 74 TGACCAAGAGCGAGTTGCCGAGCGCATCCAGAAAGACCGGGCCACCGCTCGCGCGGTGGG 133  
Db 26009 TCGTCATCGAGAGAGCGCGCGAGGCCACCGCGCCGAACTGCTCGCGCCGAAACCGCGCG 26068  
QY 134 AGGACGGCAAGAACCGGCCGACGACGCGGACCTCGTTGCCCGGTCCGCCAGGTGCTCG 193  
Db 26069 CCGAGCGGCGAGCTGCTGGTCCGAGGAGTGTGGCAGAGTGACCGTGGCCCTGATGATGT 26128  
QY 194 GCCTCGACCTCGACGAAGCCCTCGCGCGCGAGGTCTGCGCCCGCGGTCAACCGCGCGAG 253  
Db 26129 CCGCGCACAAAGAGCGCCCTCGCGACCGAGCGCGCGCGCTGCGCGCGCGACCTGCTCG 26188  
QY 254 CGACCCCAACCACTGAGACCTGGAAGAAATCGAGTGTGTCGCGACCGACCGCCCAAGCTGG 313  
Db 26189 CCCACCCGAGCTGACCGCGCGCGAGCTCGGCTACACCTCATCACCCGCGACCGCGGT 26248  
QY 314 ACGAGGA 320  
Db 26249 TCGAGCA 26255

RESULT 11  
US-09-266-965-22  
; Sequence 22, Application US/09266965  
; Patent No. 6495348  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D  
; APPLICANT: Mao, Y  
; APPLICANT: Varoglu, M  
; APPLICANT: He, M  
; APPLICANT: Sheldon, P  
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster  
; FILE REFERENCE: 600,456US1  
; CURRENT APPLICATION NUMBER: US/09/266,965  
; CURRENT FILING DATE: 1999-03-12  
; EARLIER APPLICATION NUMBER: US 08/624,447  
; EARLIER FILING DATE: 1996-08-19  
; EARLIER APPLICATION NUMBER: PCT/US94/11279  
; EARLIER FILING DATE: 1994-10-06  
; EARLIER APPLICATION NUMBER: US 08/133,963  
; EARLIER FILING DATE: 1993-10-07  
; NUMBER OF SEQ ID NOS: 145  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 22  
; LENGTH: 1224  
; TYPE: DNA  
; ORGANISM: Streptomyces lavendulae  
; US-09-266-965-22

Query Match 11.8%; Score 50.4; DB 4; Length 1224;  
Best Local Similarity 47.5%; Pred. No. 0.08;  
Matches 150; Conservative 0; Mismatches 166; Indels 0; Gaps 0;  
QY 99 CATCCAGAGACCGGGCCACCGTGGCGCGGTGGAGACCGCAAGAACCGCGCCGACGA 158  
Db 861 CGTCCAGCAGGCGCTGCGCGCGCGGTCAACCGCGAGATCGCGCGGACCAAGGT 920  
QY 159 CGCGACCTCGTTGCCCGCGTGGCCCGGTCTCGGCTCGACCTCGACGAAGCCCTCGC 218



Qy	104	AGAGGACCGGGCCACCGTCCGGCCGTGGGAGGACGGCAAGAACCGGCCGACGACGCGG	163
Db	437	GCCTCGCGCAAGGGCGAGCTGGTACGGGTCCAGCGCGGCGAGATGATTCCTCCGCGCACGGCG	496
Qy	164	ACCTCGTTGCCGCGTCCGCCAGGTGCTCGGCGCTCGACCT---CGACGAAGCCCTTCGCGC	220
Db	497	AAGTCATCGAAGGGTGGCGGCGGTCAACGAGGCGGCCATCACCGCGGAATCGGCGCGCG	556
Qy	221	CGCAGGCTCTGCGCCCGCGGCTCACCCGCCGACGACCCCAACCATGACCTTGGACGAGG	280
Db	557	TGATCCGCGAGTCCGGCGGGCGGATCGCTCGGCGGTGACCGGCAACCCACGAGTGGTGTCCG	616
Qy	281	AAATCGAGCTGTTCCGCACCGACCCCAAGCTGAGCAGGAGCATGAAGCGCGGCATCATCG	340
Db	617	ACTGCTGCTGTTACGATCCGGCGGCCAAACCCGGGGCGAGTCGACCTTGGACCCGATGATCG	676
Qy	341	CCCTAACTCTGGAGCGCGGTGAGCGGACAGAGCGCGCGCGATCGAGGAAACCAAGCGCGC	400
Db	677	CCCTGGTGGAAAGGCGCCCAACGCGGAGAGAACCCCAACGAAGTGGCGCTGGACATCTCTGC	736
Qy	401	TCATCGACCTGTTCC	415
Db	737	TGATCGGCTGACCC	751

Search completed: August 6, 2004, 13:42:31  
Job time : 58.2032 secs

```

; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9345
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9345

Query Match 11.8%; Score 50.2; DB 4; Length 1344;
Best Local Similarity 47.7%; Pred. No. 0.087;
Matches 179; Conservative 0; Mismatches 193; Indels 3; Gaps 1;

QY 44 AGCTCTCTACCGCCCGCAGCGCGCGACTGACCAAGAGCGAGTTGGCCAGCGCATCC 103
DB 1015 AGGGCTCTCCGCCAATAAGCGCGCGCGAAGCGCAGTTTCGAGAGCGTCTGCTGCAGCA 956
QY 104 AGAAGACCGGGCCACCGTCGCGGTGGAGAGCGCAAGAACCGCCCGACGACGCG 163
DB 955 GCCTGCGCAAGGCGACGTGTACGGGTTCAGGCCCGCGAGATGATTCGCGGACGCG 896
QY 164 ACCTCGTTGCCCGCGTCGCCAGGTGCTCGGCTCGACCT---CGACGAAGCCCTCGCG 220
DB 895 AAGTCATCGAAGGGTGGCGGGTCAACAGSGCGGCCATACCGCGCGAATCGGCGCG 836
QY 221 CGCAGGTCTGCCGCCCGGGTCAACCGCGCAGCACCCACCATGGACCTGGACGAGG 280
DB 835 TGATCCGCGAGTCCGCGCGCGATCGTCGCGCGGTGACCGGCACACCCAGTGGTTCG 776
QY 281 AAATCGAGTGTGTCCGCCACCGACCCCAAGCTGGACAGGACATGAAGCGCGCATCATCG 340
DB 775 ACTGGCTGTGTACGGATCGGGCCAAACCGGGCGAGTCGACCTGGACCGGATCATCG 716
QY 341 CCTAATCTCGAGCGCCGTGAGCGCGCAAGCGCGCGCGATCGAGGAAACCAAGCGGC 400
DB 715 CCTGTGTGAAGCGCGCAAGCGCGCAGAACCCCAACGAAGTGGCGCTGGACATCCTGC 656
QY 401 TCATCGACCTGTTC 415
DB 655 TGATCGGCTGACCC 641

RESULT 15
US-09-252-991A-9429
; Sequence 9429, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9429
; LENGTH: 2112
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9429

```

	Query Match	11.8%	Score 50.2;	DB 4;	Length 2112;
	Best Local Similarity	47.7%;	Pred. No. 0.088;		
	Matches 179;	Conservative	0;	Mismatches 193;	Indels 3; Gaps 1;
QY	44	ACGTCTCTCAAGCGCCGAGCGCGCCGGACTACCAAGAGCGAGTGTGCCAGGCGCATCC	103		
Db	377	AGGGCGTCTCCGCGCAATAAGCGCGCGCCGACGCGCAGTCTTCGAGAGCGTTCGTGTGCCAGCA	436		

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2004, 09:22:34 ; Search time 1649.94 Seconds  
(without alignments)  
7710.163 Million cell updates/sec

Title: US-09-855-340A-2  
Perfect score: 426  
Sequence: 1 atgcgcacacacccgggct.....acctgtccgcggagctga 426

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estnu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_estc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_est3:\*

12: gb\_est4:\*

13: gb\_est5:\*

14: gb\_est6:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rtd:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	69.8	16.4	935	29	CNS006XK
C 2	63	14.8	925	29	CNS0091P
C 3	62.8	14.7	1101	29	CNS017SY
4	62.4	14.6	582	10	BE401996 CSR003D02

5	62.4	14.6	582	13	BO607313
6	62.4	14.6	594	14	CA597375
7	62.4	14.6	620	10	BE417374
8	62.4	14.6	673	13	BO839181
9	62.4	14.6	690	14	CD869706
10	62.4	14.6	694	14	CD865484
11	62.4	14.6	872	10	BE415929
12	62.4	14.6	1035	14	CK212590
13	62.2	14.6	599	13	BO470159
14	62	14.6	553	14	CA497279
15	61.6	14.5	424	10	BE400165
16	61.2	14.4	479	13	CA025140
17	61.2	14.4	925	29	CNS0091P
18	60.8	14.3	474	10	BE498929
19	60.8	14.3	486	10	BE423849
20	60.8	14.3	590	14	CD934503
21	60.8	14.3	601	12	EM140310
22	60.8	14.3	611	10	BE401863
23	60.8	14.3	611	13	BO607482
24	60.8	14.3	633	14	CA731367
25	60.8	14.3	640	10	BE427136
26	60.8	14.3	819	13	BO807131
27	60.8	14.3	1049	14	CK184005
28	60.6	14.2	467	13	BU979539
29	60.6	14.2	491	13	CA024821
30	60.6	14.2	511	13	BU980307
31	60.6	14.2	511	13	BU981850
32	60.6	14.2	537	14	CB883553
33	60.6	14.2	550	13	BO470861
34	60.6	14.2	569	13	BO97340
35	60.6	14.2	1026	12	BG343655
36	60.2	14.1	474	10	BE416246
37	60.2	14.1	520	10	BE417736
38	60.2	14.1	619	13	BO460128
39	60	14.1	361	10	BE425095
40	59.6	14.0	1201	13	EX360624
41	59.2	13.9	1123	14	CB218083
42	59	13.8	440	14	CB883338
43	59	13.8	627	13	BU999317
44	59	13.8	748	29	CG271729
45	58.6	13.8	1097	12	BI952002

# ALIGNMENTS

RESULT 1  
CNS006XK/c 935 bp DNA linear GSS 03-JUN-1999  
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC #  
DEFINITION BAC14N09 of RPC1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
ACCESSION AL066051.1 GI:4945019  
VERSION AL066051  
KEYWORDS GSS.  
SOURCE Drosophila melanogaster (fruit fly)  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 935)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr) - Web : www.genoscope.cns.fr)  
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org/TheBDGP/Drosophila>  
melanogaster BAC library was prepared by Kazutoyo Oseegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES  
source  
1. 935  
Location/Qualifiers  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone="BACR14N09"  
/clone\_lib="RPI-98"  
/notes="end : T7"

ORIGIN  
Query Match 16.4%; Score 69.8; DB 29; Length 935;  
Best Local Similarity 30.8%; Pred. No. 1.8;  
Matches 116; Conservative 93; Mismatches 168; Indels 0; Gaps 0;

QY 3 GCGCAACACACCGGGCTGGGGCGCGGCACATGGGGCGGCATAGCTCTCACCGCCGCGA 62  
DB 920 GSGCGGCGGCGGGGCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCG 861  
QY 63 GCGCGCGCGGACTACCAAGAGCGAGTGGCCAGCGCGCATCCAGAGGACCGGGCCACCGT 122  
DB 860 GSSCCG 801  
QY 123 CGCGCGGTGGAGAGCGCAAGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 182  
DB 800 CG 741  
QY 183 CGAGGTCTCGGCTCGACCTCGACCAAGCGCTCGCGCGCGCGCGCGCGCGCGCG 242  
DB 740 GSGGCG 681  
QY 243 CACCCCG 302  
DB 680 CG 621  
QY 303 CCGCAGCTGAGGAGGAGTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 362  
DB 620 CG 561  
QY 363 GCGCGCAACAGCGCGCG 379  
DB 560 AGSGKMGASGSGRCGG 544

RESULT 2  
CNS0091P/c  
LOCUS  
DEFINITION  
Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR19D16 of RPI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Drosophila melanogaster (fruit fly)  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : [seqref@genoscope.cns.fr](mailto:seqref@genoscope.cns.fr))  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoer in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES  
source  
1. 925  
Location/Qualifiers  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone="BACR19D16"  
/clone\_lib="RPI-98"  
/notes="end : TET3"

ORIGIN  
Query Match 14.8%; Score 63; DB 29; Length 925;  
Best Local Similarity 15.1%; Pred. No. 16;  
Matches 53; Conservative 180; Mismatches 114; Indels 4; Gaps 2;

QY 3 GCGCAACACACCGGGCTGGGGCGCGGCACATGGGGCGGCATAGCTCTCACCGCCGCGA 62  
DB 900 SNSBSCG 841  
QY 63 GCGCGCGGTGGAGAGCGCAAGAGCGAGTGGCCAGCGCGCATCCAGAGGACCGGGCCACCGT 122  
DB 840 BCCMCSSSSSSCCGAGARGVKVRASGAGVRGGGSGASASHSSSSACBSSSSSSCSAS 701  
QY 123 CGCGCGGTGGAGAGCGCGCAAGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 182  
DB 780 SASSS 723  
QY 183 CGAGTCTCGGCTCGACCTCGACGAGCGCTCGCGCGCGCGCGCGCGCGCGCGCG 242  
DB 722 SSVASSMSCSS 663  
QY 243 CACCCCG 302  
DB 662 SSCSS 605  
QY 303 CCCCAAGCTGAGCGAGCATGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 353  
DB 604 SSSGGRSSGGGGGCGGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 554

RESULT 3  
CNS017SY  
LOCUS  
DEFINITION  
Drosophila melanogaster genome survey sequence Sp6 end of BAC BACN37L08 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Drosophila melanogaster (fruit fly)  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Genoscope.  
Direct Submission  
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : [seqref@genoscope.cns.fr](mailto:seqref@genoscope.cns.fr))  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a

collaboration with the European Drosophila Genome Project (EDGP) -  
<http://www.edgp.biol.ac.uk>. This Drosophila melanogaster BAC  
 library (Dros BAC) was made by Alain Billaud at CEPH (Centre  
 d'Etude du Polymorphisme Humain) with funding provided by a MRC  
 project grant. The DNA was prepared from embryos by Alain Bucheton  
 and Genevieve Payan. It has been constructed in the vector  
 pBelOBAC11.

FEATURES  
 source  
 1. .1101  
 Location/Qualifiers  
 /organism="Drosophila melanogaster"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7227"  
 /clone="BACN37L08"  
 /clone\_lib="DrosBAC"  
 /plasmid="pBelOBAC11"  
 /note="end : SP6"

## ORIGIN

Query Match 14.7%; Score 62.8; DB 29; Length 1101;  
 Best Local Similarity 16.5%; Pred. No. 17;  
 Matches 69; Conservative 184; Mismatches 165; Indels 0; Gaps 0;  
 QY 9 CACACGGGGCTGGGGCGGCGACATGGCCCATACGTCCTCACCAGCGGCGGCGCGC 68  
 DB 632 CCCACACGGGACGCGRRKGAACKAGMSGCCGSGSGCCGKKXAGVGRGRVCCA 691  
 QY 69 CGGACTGACCAAGAGCGAGCTGGCCAGGCGCATCCAGAGGACCGGCGCCAGCTCGCGCG 128  
 DB 692 GGGASCACAAACGGCCACXKACCCSSSSASSSSSCASTSSASRGWVSSCACSGSGG 751  
 QY 129 GTGGAGGACCGCAAGAACCGGCGCGACGACCGGACCTCTGTTGCCCGCTCGCCAGGT 188  
 DB 752 AGACGAGGAGGGGCGGCGCCAGGCGVCGACGACSSASCMGVSSGSSCSASGSCCG 811  
 QY 189 GCTCGGCTCGACCTCGAGAGCCCTCGCCGCGCGAGCTCGCCGCGGCTCACCCC 248  
 DB 812 VSSCAVSSASSSVNSKVASAVSCSAVASGMSAGVSSSCRSVSVSSVSSVSSV 871  
 QY 249 GCGAGCGACCCCAACCATCGACCTGGACGAGGAAATCGAGCTGTGTCGACCGACCCCA 308  
 DB 872 SSSSSSVVAAAASVSSSSASSASMAVAAAASVSVSSVSSSSSSSSSV 931  
 QY 309 GTGACGAGGACATGAACCGCGGATCATCCCTTAATCTTGGAGCGCGCTGACGCGGA 368  
 DB 932 VSASVAASASSVSSSSSSSSSSSSSVSSSVSSVSSVSSSVSSSVSSV 991  
 QY 369 CAAGGGCGGCGGATCGAGCAACCAAGCGGCTCATCGACCTGTTCCGCGCGAGCTGA 426  
 DB 992 SAAAAAASASASASAVSSSSSSSSSSSSSSSSSSSSSVSSSVSV 1049

RESULT 4  
 BE401996  
 LOCUS  
 DEFINITION  
 CSB003D02F990308 ITEC CSB Wheat Endosperm Library Triticum aestivum  
 CDNA clone CSB003D02, mRNA sequence.

ACCESSION  
 BE401996  
 VERSION  
 BE401996.1 GI:9361464  
 EST.  
 SOURCE  
 Triticum aestivum (bread wheat)  
 ORGANISM  
 Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poideae; Triticeae; Triticum.  
 1 (bases 1 to 582)

REFERENCE  
 AUTHORS  
 Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T.,  
 Cloutier, S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A.,  
 Gustafson, P., Herrmann, R.G., Holton, T., Jacquemin, J.M., Jia, J.,  
 Jouari, P., Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P.,  
 Ogihara, Y., Pecchioni, N., Qualset, C., Schuch, W., Selvaraj, G.,  
 Shariflou, M., Sorrells, M., Warburton, M. and Wenzel, G.,  
 International Triticeae EST Cooperative (ITEC): Production of  
 Expressed Sequence Tags for Species of the Triticeae

JOURNAL  
COMMENT

Unpublished (2000)  
 Contact: Appels R  
 Div. of Plant Industry, CSIRO  
 Canberra ACT 2601 AUSTRALIA  
 Tel: 61 62 465496  
 Fax: 61 62 465000  
 Email: rudi@pi.csiro.au  
 International Triticeae EST Cooperative (ITEC)  
<http://wheat.pw.usda.gov/genome>.  
 Location/Qualifiers

FEATURES  
source

1. .582  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /cultivar="Wyuna"  
 /db\_xref="taxon:4565"  
 /clone="CSB003D02"  
 /tissue\_type="endosperm"  
 /dev\_stage="8-12 days post anthesis"  
 /lab\_host="Escherichia coli SOLR"  
 /clone\_lib="ITEC CSB Wheat Endosperm Library"  
 /note="Vector: Lambda Zap/Bluescript; Site 1: XhoI;  
 Site 2: EcoRI; Plants grown in phytotron with 18C/13C  
 (day/night) 16 hour light. M13 Reverse sequencing primer  
 used. 1.0 Kbp average insert size."

## ORIGIN

Query Match 14.6%; Score 62.4; DB 10; Length 582;  
 Best Local Similarity 50.9%; Pred. No. 19;  
 Matches 175; Conservative 0; Mismatches 166; Indels 3; Gaps 1;  
 QY 67 GCGGACTGACCAAGACCGAGTGGCCAGGCGGATCCAGAGG---ACCGGGCCACCGTC 123  
 DB 78 GCGCGAGAGAGGAGAGAGCGGTGACCGGTCCACCAAGCGCGCTCCAGATTCCCCGTC 137  
 QY 124 GCGCGTGGGAGGACGCAAGAACCGGCGCGACGACCGGACCTGTTGCCCGGCTCGCC 183  
 DB 138 GCGCGCATCGGCGCTTCTCAGAAGGCGCGTACGCGAGCGGCTCGCTCGGCGCC 197  
 QY 184 CAGTGTCTCGGCTTCGACCTCGACGAAGCCCTCGCGCGCGAGGTCTGCGCGCGCGGTC 243  
 DB 198 CCGCTCTACCTCGCGCGCGCTCTCTGAGTACCTCGCGCGCGAGGTCTGAGCTCGCGCG 257  
 QY 244 ACCCGCGCAGGCGCCCAACCATGAGCTGGAGGAGAAATCAGCTGTCTCGCACCAGC 303  
 DB 258 AACCGCGCAAGGACAACAAGAGACCCCGCATGTGCGCGCGCACCTGCTCTCGGCATC 317  
 QY 304 CCCAAGCTGACGAGGACATGAAGCGCGCATCATCGCCCTTAATCTTGGAGCGCGGTGAG 363  
 DB 318 GCGAACGACGAGGAGCTCGGCAAGCTCTCGCGGCGTACCATCGGCGCAGCGCGCGTG 377  
 QY 364 GCGGACAGGCGGCGGCGATCGAGGAAACCAAGCGGCTCATCGA 407  
 DB 378 CTCGCCAACATCAACCCCGCTGCTCTCCCAAGAGGCGCTCGA 421

## RESULT 5

BQ607313

LOCUS

DEFINITION

BQ607313

ACCESSION

BQ607313.1

VERSION

BQ607313

KEYWORDS

EST.

SOURCE

Triticum aestivum

ORGANISM

Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poideae; Triticeae; Triticum.  
 1 (bases 1 to 582)

REFERENCE

AUTHORS

Clarke, B., Lambrecht, M. and Rhee, S.Y.

TITLE

Arabidopsis genomic information for interpreting wheat EST  
 sequences

JOURNAL

Funct. Integr. Genomics 3 (1-2), 33-38 (2003)

MEDLINE  
PUBMED  
COMMENT

22478026  
12590341  
Contact: Lambrecht M  
The Arabidopsis Information Resource  
Carnegie Institution of Washington, Dept. of Plant Biology  
260 Panama Street, Stanford, CA 94305, USA  
Tel: 1 650 325 1521 x 251  
Fax: 1 650 325 3748  
Email: rhes@acoma.stanford.edu.  
Location/Qualifiers

FEATURES  
source

1..582  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="Wyuna"  
/db\_xref="taxon:4565"  
/tissue\_type="endosperm"  
/dev\_stage="developing endosperm tissue 8, 10 and 12 DPA (days post anthesis)"  
/clone\_lib="wheat EST endosperm library"

ORIGIN

Query Match 14.6%; Score 62.4; DB 13; Length 582;  
Best Local Similarity 50.9%; Pred. No. 19;  
Matches 175; Conservative 0; Mismatches 166; Indels 3; Gaps 1;  
QY 67 GCCGACTGACCAAGAGCGAGTTGGCCAGCGCGATCCAGAAGG---ACCGGCCACCGTC 123  
Db 78 GCGGAGAGAGGAAGAAGCGGTGACCGGTCCACCAAGCGCGCTCCAGTTCCCGCTC 137  
QY 124 GCGCGGTGGAGACGCGAAGACCGCGCCGCGACGCGGACCTGTTCCCGCGTCGCC 183  
Db 138 GCGCGCATCGGCGCTTCTCAAGAGCGCGGTACGCGAGCGGTGCGTCCCGCGCC 197  
QY 184 CAGTGCTCGGCTCGACCTCGACGAGCCCTGCGCGCGAGGTGTCGCCCGCGCGTC 243  
Db 198 CCGGTCTACTCGCGCGCTCTCGAGTACCTCGCGCGCGAGGTGTCGAGCTCGCGCG 257  
QY 244 ACCCGCGAGCAGCCCAACCATGAGCTGGACGAGGAATCGAGTGTGCGCGCGCGAC 303  
Db 258 AACCGCGCAAGACAAACAGAGACCCGCGATCGTCCGCGCGACCTGCTCTCGCCATC 317  
QY 304 CCCAAGCTGACGAGGACATGAAGCGCGCGATCATCGCCCTAATCTGTGAGCGCGTGAG 363  
Db 318 CGCAACGACGAGGCTCGCAAGTCTGCGCGGCTCACCATCGCGCACGCGCGCGTG 377  
QY 364 CGGACAAAGCGCGCGGATCGAGGAACCAAGCGGCTCATCGA 407  
Db 378 CTCCCAACATCAACCCCGTGTGTCTCCCAAGAGGCGCCCTCGA 421

RESULT 6  
CA597375

LOCUS  
DEFINITION  
wpaic.pk016.n10 wpaic Triticum aestivum cDNA clone wpaic.pk016.n10  
5' end, mRNA sequence.

ACCESSION  
CA597375

VERSION  
CA597375.1 GI:25150171

KEYWORDS  
EST

SOURCE  
Triticum aestivum (bread wheat)

ORGANISM  
Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poaceae; Triticeae; Triticum.

REFERENCE  
1 (bases 1 to 594)

AUTHORS  
Tingey, S.V., Moore, G., Griffiths, S., Powell, W., Wolters, P.,

Dolan, M., Hainey, C., Miao, G., Caraher, N. and Hanafey, M.K.

DuPont Wheat cDNA Sequence in collaboration with the John Innes

Center 1

Unpublished (2002)

CONTACT: Scott V. Tingey

Crop Genetics

E. I. DuPont de Nemours and Company

1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA

Tel: 302-631-2602  
Fax: 302-631-2607  
Email: Scott.V.Tingey@USA.dupont.com

Seq primer: M13.

FEATURES  
source

Location/Qualifiers  
1..594  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/db\_xref="taxon:4565"  
/clone\_lib="wpaic.pk016.n10"  
/tissue\_type="anthers"  
/lab\_host="DH10B"  
/clone\_lib="wpaic"  
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:  
XhoI; Wheat (Triticum aestivum) pre-meiotic anthers JIC"

ORIGIN

Query Match 14.6%; Score 62.4; DB 14; Length 594;  
Best Local Similarity 50.9%; Pred. No. 19;  
Matches 175; Conservative 0; Mismatches 166; Indels 3; Gaps 1;  
QY 67 GCCGACTGACCAAGAGCGAGTTGGCCAGCGCGATCCAGAAGG---ACCGGCCACCGTC 123  
Db 72 GCGGAGAGAGGAAGAAGCGGTGACCGGTCCCAAGCGCGCTCCAGTTCCCGCTC 131  
QY 124 GCGCGGTGGAGACGCGAAGACCGCGCCGCGACGCGGACCTGTTGCCCGCGTCGCC 183  
Db 132 GCGCGCATCGGCGCTTCTCAAGAGAGGCGCGTACGCGAGCGCGTCCGCTCCGCGCC 191  
QY 184 CAGTGCTCGGCTCGACCTCGACGAGCCCTGCGCGCGAGGTGTCGCCCGCGCGTC 243  
Db 192 CCGGTCTACTCGCGCGCTCTCGAGTACCTCGCGCGAGGTGTCGAGCTCGCGCG 251  
QY 244 ACCCGCGAGCAGCCCAACCATGAGCTGGACGAGGAATCGAGTGTGCGCGCGCGAC 303  
Db 252 AACCGCGCAAGACAAACAGAGAGCGCGATCGTCCGCGCGCACCTGCTCTCGCCATC 311  
QY 304 CCCAAGCTGACGAGGACATGAAGCGCGCGATCATCGCCCTAATCTGTGAGCGCGTGAG 363  
Db 312 CGCAACGACGAGGCTCGCAAGTCTGCGCGGCTTACCATCGCGCACGCGCGCGTG 371  
QY 364 CGGACAAAGCGCGCGGATCGAGGAACCAAGCGGCTCATCGA 407  
Db 372 CTGCCCAACATCAACCCCGTGTGTCTCCCAAGAGGCGCCCTCGA 415

RESULT 7  
BE417374

LOCUS  
DEFINITION  
MUG020.D01R990620 ITEC MUG Wheat Spikelet Library Triticum aestivum  
cDNA clone MUG020.D01, mRNA sequence.

ACCESSION  
BE417374

VERSION  
BE417374.1 GI:9415220

KEYWORDS  
EST

SOURCE  
Triticum aestivum (bread wheat)

ORGANISM  
Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poaceae; Triticeae; Triticum.

REFERENCE  
1 (bases 1 to 620)

AUTHORS  
Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T.,

Cloutier, S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A.,

Gustafson, P., Herrmann, R.G., Holt, T., Jacquemin, J.M., Jia, J.,

Joudrier, P., Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P.,

Ogihara, Y., Pecchioni, N., Qualset, C., Schuch, W., Selvaraj, G.,

Sharflou, M., Sorrells, M., Warburton, M. and Wenzel, G.

International Triticeae EST Cooperative (ITEC): Production of

Expressed Sequence Tags for Species of the Triticeae

Unpublished (2000)

CONTACT: Ogihara Y

Kihara Institute for Biological Research, Yokohama City University

Maoka-cho 641-12, Totsuka-ku, Yokohama 244-0813, JAPAN

Tel: 81 45 820 1903

Fax: 81 45 820 1901  
 Email: ogihara@yokohama-cu.ac.jp  
 International Triticeae EST Cooperative (ITEC)  
 http://wheat.pw.usda.gov/genome.

## FEATURES

Location/Qualifiers  
 1..620  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /cultivar="Norin 26"  
 /db\_xref="taxon:4565"  
 /clone="WUG020.D01"  
 /tissue\_type="young spikelets"  
 /dev\_stage="feekes' scale 6-7"  
 /clone\_lib="ITEC MUG Wheat Spikelet Library"  
 /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; M13 Reverse sequencing primer used. 1.2 kbp average insert size."

## ORIGIN

Query Match 14.6%; Score 62.4; DB 10; Length 620;  
 Best Local Similarity 50.9%; Pred. No. 19;  
 Matches 175; Conservative 0; Mismatches 166; Indels 3; Gaps 1;

QY 67 GCGGACTACCAAGAGCGAGTTGGCCAGGCGCATCCAGAGG---ACCGGCCACCGTC 123  
 Db 5 GCGGAGAGAGAGAGCGGTGACCCGGTCCACCAAGCGCGCTCCAGTTCCCGTC 64

QY 124 GCGCGTGGAGAGCGGCAAGAACCGGCCCGACGACGCGGACCTGTTGCCCGCGTCGCC 183  
 Db 65 GCGCGCATCGGCGCTTCTCTCAAGAAGGCGGTACGCGACGCGGTGCGGTCCCGCGCC 124

QY 184 CAGGTCTCGGCTCGACCTCGACGAAGCCCTCGCGCGCAGGTCGCGCCCGCGTC 243  
 Db 125 CCGGTCTACTCGCGCGCGCTCTCGAGTACCTTCGCGCGCGAGGTGCTGGAGTCCCGGC 184

QY 244 ACCCGCCACGACCCCAACCATGACCTGACGAGGAATCGAGCTGTCGCCACCGAC 303  
 Db 185 AACCGCGCAAGACACCAAGAGACCCGACATGTCGCGCGCACCTGCTCTCGCCATC 244

QY 304 CCGAGCTGGAGAGACATGAAGCGCGCATCATCGCCCTTAATCTCGAGGCGCGTGAG 363  
 Db 245 CGCAACGACGAGAGCTCGGCAAGTGTCTCGCGCGCTCACCATCGCGACGCGCGGTG 304

QY 364 CGCGACAAGCGCGCGCATCGAGGAACCAAGCGGCTCATCGA 407  
 Db 305 CTCCTCCACATCAACCCCGTGTCTCTCCCAAGAGGCGCTCGA 348

RESULT 8  
 BQ839181  
 LOCUS  
 DEFINITION  
 WHE4163\_C06\_E11ZS Wheat CS whole plant cDNA library  
 aestivum cDNA clone WHE4163\_C06\_E11, mRNA sequence.

ACCESSION  
 BQ839181  
 VERSION  
 BQ839181.1 GI:22143503  
 EST.  
 Triticum aestivum (bread wheat)  
 Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poideae; Triticeae; Triticum.  
 1 (bases 1 to 673)  
 Anderson, O.D., Akhunov, E., Chao, S., Crossman, C., Deal, K.,  
 Dvorak, J., Izzo, G.R., Rausch, C.J., Wilson, C. and Woo, J.  
 The structure and function of the expressed portion of the wheat  
 genomes - Chinese Spring whole plant cDNA library  
 Unpublished (2002)  
 Contact: Olin Anderson  
 US Department of Agriculture, Agriculture Research Service, Pacific  
 West Area, Western Regional Research Center  
 800 Buchanan Street, Albany, CA 94710, USA  
 Tel: 5105595773  
 Fax: 5105595818

Email: andersn@pw.usda.gov  
 Sequences have been trimmed  
 quality sequence with phred score less than 20  
 Seq primer: SK primer.

## FEATURES

Location/Qualifiers  
 1..673  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /cultivar="Chinese Spring"  
 /db\_xref="taxon:4565"  
 /clone="WHE4163\_C06\_E11"  
 /tissue\_type="Roots, leaves, crown, stem and sheath"  
 /dev\_stage="Adult"  
 /lab\_hosts="E. coli SGRP"  
 /clone\_lib="Wheat CS whole plant cDNA library"  
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid  
 pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; Plant  
 tissues from wheat cv. CS grown to full tillering stage in  
 greenhouse were collected at University of California,  
 Davis (Jan Dvorak's lab, K. Deal and E. Akhunov). Total  
 RNA was prepared from leaves (young leaf and third leaf),  
 whole roots, crown, stem and sheath tissues, and then  
 equal quantities of RNA were pooled from the these  
 samples. PolyA was purified from the pooled RNA, a cDNA  
 library was made, and the cDNA clones were in vivo excised  
 to give pBluescript SK(-) phagemids in J. Dvorak's lab (E.  
 Akhunov, J. Dvorak) at the University of California,  
 Davis. Colony plating, plasmid DNA preparations and DNA  
 sequencing were performed in the OD Anderson lab (all  
 other authors)."

## ORIGIN

Query Match 14.6%; Score 62.4; DB 13; Length 673;  
 Best Local Similarity 50.9%; Pred. No. 19;  
 Matches 175; Conservative 0; Mismatches 166; Indels 3; Gaps 1;

QY 67 GCGGACTACCAAGAGCGAGTTGGCCAGGCGCATCCAGAGG---ACCGGCCACCGTC 123  
 Db 65 GCGGAGAGAGAGAGGCGGTGACCCGGTCCACCAAGCGCGCTCCAGTTCCCGTC 124

QY 124 GCGCGTGGAGAGCGGCAAGAACCGGCCCGACGACGCGGACCTGTTGCCCGCGTCGCC 183  
 Db 125 GCGCGCATCGGCGCGCTTCTCTCAAGAAGGCGGTACGCGCGCGCTGCTGGAGTCCCGGC 184

QY 184 CAGGTGCTCGGCGCTCGACCTCGAGGAAGCCCTCGCGCGCGAGGTCTCGCGCCCGCGTC 243  
 Db 185 CCGGTCTACTCGCGCGCGTCTCTCGAGTACCTCGCGCGCGAGGTGCTGGAGTCCCGGC 244

QY 244 ACCCGCCACGACCCCAACCATGACCTGACGAGGAATCGAGCTGTCGCCACCGAC 303  
 Db 245 AACCGCGCAAGACACCAAGAGACCCCGCATGTCGCGCGCGCTCTCTCGCCATC 304

QY 304 CCCAAGCTGGACGAGGACATGAAGCGCGCATCATCGCCCTTAATCTCTGGAGCGCGTGAG 363  
 Db 305 CGCAACGACGAGGCTCGGCAAGTGTCTCGCGCGCTCACCATCGCGCACGCGCGGTG 364

QY 364 CGCGACAAGCGCGCGGATCGAGGAACCAAGCGGCTCATCGA 407  
 Db 365 CTCCTCCACATCAACCCCGTGTCTCTCCCAAGAGGCGCTCGA 408

RESULT 9  
 CD869706  
 LOCUS  
 DEFINITION  
 AZ02.112G03F001120 AZ02 Triticum aestivum cDNA clone AZ02112G03,  
 mRNA sequence.

ACCESSION  
 CD869706  
 VERSION  
 CD869706.1 GI:32553522  
 KEYWORDS  
 EST.  
 Triticum aestivum (bread wheat)  
 Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Poideae; Triticeae; Triticum.

REFERENCE  
1 (bases 1 to 690)  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Genoplante.  
Unpublished (2003)  
Contact: Genoplante  
Genoplante  
93, rue Henri Rochefort 91025 EVRY CEDEX France  
Tel: 33 1 69 47 54 00  
Fax: 33 1 69 47 54 10  
This sequence has been generated in the framework of the french  
plant genomics programme 'Genoplante' (<http://www.genoplante.com>)  
and <http://genoplante-info.infobiogen.fr>.

FEATURES  
source

Location/Qualifiers  
1..690  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="recital"  
/db\_xref="taxon:4565"  
/clone="AZO2112603"  
/tissue\_type="root"  
/clone\_lib="AZO2"

ORIGIN

Query Match 14.6%; Score 62.4; DB 14; Length 690;  
Best Local Similarity 50.9%; Pred. No. 19;  
Matches 175; Conservative 0; Mismatches 166; Indels 3; Gaps 1;  
QY 67 GCCGACTGACCAAGAGCGAGTTGGCCAGGCGGATCCAGAAGG---ACCGGCGCACCGTC 123  
DB 85 GCGCGAGAGAGAGAGAGCGGTGACCCCGGTCCACCAAGCGCGGCTCCAGTTCCTCCCGTC 144  
QY 124 GCGCGGTGGAGAGACGCGCAAGAACCGCGCCGCGAGCGGACCTGCTTCCCGCGTCGCG 183  
DB 145 GCGCGATCGGGCGCTTCTCAGAGAGGCGCGTACGCGAGCGGTGCGTTCGCGCGCC 204  
QY 184 CAGTGCTCGGCTCGACCTCGACGAGCCCTGCGCGCGCGAGGTGTGCGCCCGCGCGTC 243  
DB 205 CCGGTCTACCTCGCGCGCGTCTCGAGTACCTCGCGCGCGAGGTGTGAGCTCGCGCGC 264  
QY 244 ACCCGCGAGCGACCCACCACTGAGTGGAGGAAATCGAGTGTTCGCGACCGAC 303  
DB 265 ACGCGCGCAAGAGCAACAGAGACCCGCGATCGTCCGCGCGACCTGCTTCTCGCGATC 324  
QY 304 CCAAGCTGAGGAGGACATGAAGCGCGGCGATCATCGCCCTAATCTCGAGCGCGTGAG 363  
DB 325 CGCAACGACGAGGATCGGCAAGTGTCTCGCGCGGCTCACCATCGCGCACGCGCGGTG 384  
QY 364 CGGCAAGGCGCGCGGATCGAGGAAACCAAGCGGCTCATCGA 407  
DB 385 CTCCCAACATCAACCCCGTGTCTCTCCCAAGAGGCGCTCGA 428

RESULT 10  
CD865484

LOCUS  
DEFINITION  
AZO2.101A18F010111 AZO2 Triticum aestivum cDNA clone AZO2101A18,  
mRNA sequence.

CD865484.1 GI:32549300

ACCESSION  
VERSION  
KEYWORDS  
SOURCE

ORGANISM  
Triticum aestivum (bread wheat)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Triticum.

REFERENCE  
1 (bases 1 to 694)

Genoplante.  
Unpublished (2003)  
Contact: Genoplante  
Genoplante  
93, rue Henri Rochefort 91025 EVRY CEDEX France

Tel: 33 1 69 47 54 00  
Fax: 33 1 69 47 54 10

This sequence has been generated in the framework of the french  
plant genomics programme 'Genoplante' (<http://www.genoplante.com>)  
and <http://genoplante-info.infobiogen.fr>.

FEATURES  
source

Location/Qualifiers  
1..694  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="recital"  
/db\_xref="taxon:4565"  
/clone="AZO2101A18"  
/tissue\_type="root"  
/clone\_lib="AZO2"

ORIGIN

Query Match 14.6%; Score 62.4; DB 14; Length 694;  
Best Local Similarity 50.9%; Pred. No. 19;  
Matches 175; Conservative 0; Mismatches 166; Indels 3; Gaps 1;  
QY 67 GCCGACTGACCAAGAGCGAGTTGGCCAGGCGGATCCAGAAGG---ACCGGCGCACCGTC 123  
DB 76 GCGCGAGAGAGAGAGAGCGGTGACCCCGGTCCACCAAGCGCGGCTCCAGTTCCTCCCGTC 135  
QY 124 GCGCGGTGGAGAGACGCGCAAGAACCGCGCCGCGAGCGGACCTGCTTCCCGCGTCGCG 183  
DB 136 GCGCGATCGGGCGCTTCTCAGAGAGGCGCGTACGCGAGCGGTGCGTTCGCGCGCC 195  
QY 184 CAGTGCTCGGCTCGACCTCGACGAGCCCTGCGCGCGCGAGGTGTGCGCCCGCGCGTC 243  
DB 196 CCGGTCTACCTCGCGCGCGTCTCGAGTACCTCGCGCGCGAGGTGTGAGCTCGCGCGC 255  
QY 244 ACCCGCGAGCGACCCACCACTGAGTGGAGGAAATCGAGTGTTCGCGACCGAC 303  
DB 256 ACGCGCGCAAGAGCAACAGAGACCCGCGATCGTCCGCGCGACCTGCTTCTCGCGATC 315  
QY 304 CCAAGCTGAGGAGGACATGAAGCGCGGCGATCATCGCCCTAATCTCGAGCGCGTGAG 363  
DB 316 CGCAACGACGAGGATCGGCAAGTGTCTCGCGCGGCTCACCATCGCGCACGCGCGGTG 375  
QY 364 CGGCAAGGCGCGCGGATCGAGGAAACCAAGCGGCTCATCGA 407  
DB 376 CTCCCAACATCAACCCCGTGTCTCTCCCAAGAGGCGCTCGA 419

RESULT 11  
BE415929

LOCUS

DEFINITION  
MUG002.E06R990520 ITEC MUG Wheat Spikelet Library Triticum aestivum  
cDNA clone MUG002.E06, mRNA sequence.

BE415929.1 GI:9413775

ACCESSION  
VERSION  
KEYWORDS  
SOURCE

ORGANISM  
Triticum aestivum (bread wheat)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Triticum.

REFERENCE  
1 (bases 1 to 872)

Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier, S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P., Herrmann, R.G., Holton, T., Jacquemin, J.M., Jia, J., Joudrier, P., Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P., Ogihara, Y., Pecchioni, N., Qualset, C., Schuch, W., Selvaraj, G., Shariflou, M., Sorrells, M., Warburton, M. and Wenzel, G.  
International Triticale EST Cooperative (ITEC): Production of  
Expressed Sequence Tags for Species of the Triticeae  
Unpublished (2000)  
Contact: Ogihara Y  
Kihara Institute for Biological Research, Yokohama City University  
Maoka-cho 641-12, Totsuka-Ku, Yokohama 244-0813, JAPAN  
Tel: 81 45 820 1903  
Fax: 81 45 820 1901

Email: ogihara@okohama-cu.ac.jp  
International Triticeae EST Cooperative (ITEC)  
http://wheat.pw.usda.gov/genome.  
Location/Qualifiers

FEATURES

1..872  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="Norin 26"  
/db\_xref="taxon:4565"  
/clone="MUG002.E06"  
/tissue\_type="young spikelets"  
/dev\_stages="Feekes' scale 6-7"  
/clone\_lib="ITEC MUG Wheat Spikelet Library"  
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; M3 Reverse sequencing primer used. 1.2 kbp average insert size."

ORIGIN

Query Match 14.6%; Score 62.4; DB 10; Length 872;  
Best Local Similarity 50.9%; Pred. No. 19;  
Matches 175; Conservative 0; Mismatches 166; Indels 3; Gaps 1;  
  
QY 67 GCGGACTGACCAAGAGCGAGTTGCCAGGCGCATCCAGAGG---ACGGGCGACCGTC 123  
Db 81 GCGGAGAGAGGAGGAGCGGTGACCGGTCCTCAAGGCGCGCTCAGTTCCCGTC 140  
  
QY 124 GCGGCTGGGAGAGCGGAAGAACCGGCGGACGAGCGGACCTCGTTGCCGCGTCGCC 183  
Db 141 GCGGCGATCGGCGGCTTCTCAAGAGGCGCGCTACGCGAGCGGTGCGGTCCGCGCC 200  
  
QY 184 CAGGTGCTGGGCTGACCTCGACGAGCGCTCGCGGCGGAGGTCTGCGCCCGCGTC 243  
Db 201 CCGCTTACTCTGCGCGCGCTCTCGAGTACCTCGCGCGGAGGTCTGAGTCTCGCGCGC 260  
  
QY 244 ACCGCGCGAGCGACCCACCACTGACCTGAGCGAGGAAATCGAGTGTCTCGCACCGAC 303  
Db 261 AACGCGCGCAAGAGCAACAGAGAGCGCGATCTGCGCGCGACCTGCTCTCGCCATC 320  
  
QY 304 CCGAGCTGCGAGAGCATGAGGCGCGCATATCGCCCTAATCTCTGAGGCGCGGTAG 363  
Db 321 CGCAACGACGAGGCTCGGCAAGTGTCTGCGCGCGTCCACATCGCGCACGCGCGCGTG 380  
  
QY 364 CGCGACAGGCGGCGCGATCGAGGAACCAAGCGGCTCATCGA 407  
Db 381 CTGCCCCAATCAACCCCGTGTCTCCCAAGAGGCGCGCTCGA 424

RESULT 12

CK212590/c  
LOCUS CK212590.1  
DEFINITION FGAS024469 Triticum aestivum FGAS: Library 6 CAP GATE 1 Triticum  
aestivum cDNA, mRNA sequence.

ACCESSION

VERSION CK212590  
KEYWORDS EST.  
SOURCE Triticum aestivum (bread wheat)

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Triticum.

REFERENCE

AUTHORS Allard, F., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,  
Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Laroche, A.,  
Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilsson, D.,  
Penniket, C., Roach, J.B. and Sarhan, F.  
Contact: Wm L Crosby

TITLE

Functional Genomics of Abiotic Stress in Wheat and Canola Crops  
UNPUBLISHED (2003)  
Contact: Wm L Crosby  
Bioinformatics  
University of Saskatchewan, Department of Computer Science  
1C101 Engineering Building, 57 Campus Drive, Saskatoon,  
Saskatchewan, S7N 5A9, Canada  
Tel: 306 966 1769

Fax: 306 966 2033  
Email: fgas.ests@cs.usask.ca

This sequence is the direct result of the Base calling software  
Phred (default parameters). It is the raw base calls. To aid in the  
identification of the high quality insert the software Lucy  
(default parameters) has been run on this sequence. Lucy identified  
the region [21,782].  
Plate: L6B006 row: E column: 04.

FEATURES

Location/Qualifiers  
1..1035  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/db\_xref="taxon:4565"

/note="Organ: Crown and leaf; Vector: pCMV.SPORT6; Crown  
(50%) and leaf (50%) tissues from wheat cultivar Norstar  
after short exposure times to low temperature in the light  
and in the dark. 12 mRNA populations were combined before  
constructing the library. The first 6 populations: After 7  
days of growth at 20C from wheat cultivar Norstar after  
short exposure times to low temperature in the light and  
in the dark. 12 mRNA populations were combined before  
constructing the library. The first 6 populations: After 7  
days of growth at 20, wheat plants were transferred to 4C  
in the light. 1cm crown sections and green leaf tissue were  
separately harvested after 1, 3, and 6 hours of low  
temperature exposure. The last 6 populations: After 7 days  
of growth at 20C, wheat plants were transferred to 4C in  
the dark. 1cm crown sections and green leaf tissue were  
separately harvested after 1, 3, and 6 hours of low  
temperature exposure. First strand synthesis in this  
library was done in the presence of methylated dCTP  
thereby protecting from internal cleavage with NotI. In  
addition, this library used a primer for second strand  
synthesis that annealed to an artificial sequence (RNA  
oligo) added before first strand synthesis. Therefore when  
sequences from EST generated from this library will be  
masked for vector and adaptor sequences, an additional  
masking step will have to be included to mask this RNA  
oligo that is common to all clones (sequence  
CGACTGAGCAGGAGGACATGATGACTGAGTGAAGGTAGAAA)."

ORIGIN

Query Match 14.6%; Score 62.4; DB 14; Length 1035;  
Best Local Similarity 50.9%; Pred. No. 19;  
Matches 175; Conservative 0; Mismatches 166; Indels 3; Gaps 1;  
  
QY 67 GCGGACTGACCAAGAGCGAGTTGCCAGGCGCATCCAGAGG---ACGGGCGACCGTC 123  
Db 718 GCGGAGAGAGGAGGAGCGGTGACCGGTCCTCAAGGCGCGCTCAGTTCCCGTC 659  
  
QY 124 GCGGCTGGGAGAGCGGAAGAACCGGCGGACGAGCGGACCTCGTTGCCGCGTCGCC 183  
Db 658 GCGGCGATCGGCGCGCTTCTCAAGAGGCGCGTACGCGAGCGCGTCTCGCGCGCC 599  
  
QY 184 CAGGTGCTGGGCTTCAAGAGCGCGCTCGCGCGCGAGGTCTGCGCGCGCGCGTC 243  
Db 598 CCGCTTACTCTGCGCGCGCTCTCGAGTACCTCGCGCGGAGGTCTGAGTCTCGCGCGC 539  
  
QY 244 ACCGCGCGAGCGACCCACCACTGACCTGAGCGAGGAAATCGAGTGTCTCGCACCGAC 303  
Db 538 AACGCGCGCAAGAGCAACAGAGAGCGCGATCTGCGCGCGACCTGCTCTCGCCATC 479  
  
QY 304 CCGAGCTGCGAGGAGCATGAGCGGCGCATATCGCCCTAATCTCTGAGGCGCGGTAG 363  
Db 478 CGCAACGACGAGGAGTCTGGCAAGTGTCTGCGCGCGTCCACATCGCGCACGCGCGCGTG 419  
  
QY 364 CGCGACAGGCGGCGGCGATCGAGGAACCAAGCGGCTCATCGA 407  
Db 418 CTGCCCCAATCAACCCCGTGTCTCCCAAGAGGCGCGCTCGA 375

RESULT 13



Db 124 CCTCGCGACCTCGTCAACGCTGTTCCCTGCGCCCGCGCTCAAGGACGCCGCCGCTGCGC 183  
 QY 264 CATGACCTGACGAGGAAATCGAGCTGTCGCGACCGACCCCAAGCTGGACGAGGACAT 323  
 Db 184 CATCTCCCTTCGAGGACTCGTCGACGGGATCAGCGGCAACCTCTTCGACGCTAGGT 243  
 QY 324 GAAGCGGCGCATCATGCGCCCTAATCCTGAGGCGCGTGGAGCGGACAAAGCGGCGGCGAT 383  
 Db 244 CAAGCCCTACTTCAATTGCGCGCTGCGCGCGTGGCGAAGGCGGACCGGTTCTGTGTGA 303  
 QY 384 CG 385  
 Db 304 CG 305

RESULT 15  
 BE400165 424 bp mRNA linear EST 21-JUL-2000  
 LOCUS AWB011.D10F000328 ITEC AWB Wheat Meiotic Stage Library Triticum  
 DEFINITION aestivum cDNA clone AWB011.D10, mRNA sequence.

ACCESSION BE400165  
 VERSION BE400165.1 GI:9359633  
 KEYWORDS EST.  
 SOURCE Triticum aestivum (bread wheat)  
 ORGANISM Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poideae; Triticeae; Triticum.  
 1 (bases 1 to 424)

REFERENCE  
 AUTHORS Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T.,  
 Cloutier, S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A.,  
 Gustafson, P., Herrmann, R.G., Holton, T., Jacquemin, J.M., Jia, J.,  
 Joudrier, P., Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P.,  
 Ogihara, Y., Pecchioni, N., Qualset, C., Schuch, W., Selvaraj, G.,  
 Shariflou, M., Sorrells, M., Warburton, M. and Wenzel, G.  
 International Triticeae EST Cooperative (ITEC): Production of  
 Expressed Sequence Tags for Species of the Triticeae  
 Unpublished (2000)

JOURNAL  
 COMMENT Contact: Langridge P  
 Special Research Centre, Basic and Applied Plant Molecular Biology  
 Waite Campus, University of Adelaide, Glen Osmond SA 5064 AUSTRALIA  
 Tel: 61 8 8303 7368  
 Fax: 61 8 8303 7102  
 Email: plangridge@waite.adelaide.edu.au  
 International Triticeae EST Cooperative (ITEC)  
 http://wheat.pw.usda.gov/genome.  
 Location/Qualifiers

FEATURES  
 source  
 1..424  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /cultivar="Chinese Spring"  
 /db\_xref="taxon:4565"  
 /clone="AWB011.D10"  
 /tissue\_type="excised florets"  
 /dev\_stage="meiotic stage no later than metaphase I"  
 /clone\_lib="ITEC AWB Wheat Meiotic Stage Library"  
 /notes="Vector: pSport 1 (Life Technologies cat. no.  
 18248-013); Site 1: SalI; Library constructed in pSport 1.  
 Directionally cloned using the Superscript Plasmid System  
 for cDNA synthesis and plasmid cloning. M13 Reverse  
 sequencing primer used to obtain 5' sequence data. 1.4  
 Kbp average insert size."

## ORIGIN

Query Match 14.5%; Score 61.6; DB 10; Length 424;  
 Best Local Similarity 52.3%; Pred.No. 23;  
 Matches 161; Conservative 0; Mismatches 144; Indels 3; Gaps 1;  
 QY 67 GCGGACTGACGAGCGAGTTGGCCAGCGCGCATCCAGAAGG---ACCGGCGCACCGTC 123  
 Db 91 GCGGAGAGGAGGAGGCGGTGACCCCGTCCACCAAGCGCGCTCCAGTTCGCCGTC 150

QY 124 GCGCGTGGAGACGGCAAGAACCGCGCGACGCGGACCTGTTCCCGCGCTGCGC 183  
 Db 151 GCGCGATCGGGCGCTTCTCAAGAGGGCGCTACGCGAGCGCTCGGCTCCGCGGC 210  
 QY 184 CAGGTGTCGCGCCTCGACCTCGACGAAAGCCCTCGCCCGCGCAGGTCTGGCGCCCGCGCTC 243  
 Db 211 CCGGTCTACCTCGCGCGCTCTCGAGTACTCTCGCGCGGAGGTGCTGGAGTTCGCGCGC 270  
 QY 244 ACCCGCGCAGGACCCCAACCATGGACCTGGAGAGGAAATCGAGTGTTCGCGACCGAC 303  
 Db 271 AACGCGCCCAAGAACAAACAAAGAGACCCCGCATCTGCGCGCGCACCTGTCTCTCGCATC 330  
 QY 304 CCCAAGCTGGACGAGGACATGAAGCGCGCGCATCATGCCCCCTAATCTTGGAGCGCCCTGAG 363  
 Db 331 CGCAACGACGAGGCTCGGCAGCTGCTCGCGGGGTACCATCGCGCAGCGGCGGTG 390  
 QY 364 CGGACAA 371  
 Db 391 CTCGCCAA 398

Search completed: August 6, 2004, 13:39:02  
 Job time : 1653.94 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: August 6, 2004, 08:50:04 ; Search time 156.122 Seconds  
(without alignments)  
9439.175 Million cell updates/sec

Title: US-09-855-340A-3  
Perfect score: 34  
Sequence: 1 ccccgtaggggttcattccatccagtcacccg 34

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues  
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:

- 1: gb\_ba:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vii:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vii:\*
- 30: em\_htg\_hum:\*
- 31: em\_htg\_inv:\*
- 32: em\_htg\_other:\*
- 33: em\_htg\_mus:\*
- 34: em\_htg\_pin:\*
- 35: em\_htg\_rod:\*
- 36: em\_htg\_mam:\*
- 37: em\_htg\_vrt:\*
- 38: em\_sy:\*
- 39: em\_htgo\_hum:\*
- 40: em\_htgo\_mus:\*
- 41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	34	100.0	34	6	AX338972	AX338972 Sequence
2	34	100.0	145	1	AY150033	AY150033 Micromono
3	34	100.0	247	6	AX338975	AX338975 Sequence
4	34	100.0	2025	7	AY150027	AY150027 Bacteriop
5	33	97.1	143	1	AY150032	AY150032 Micromono
6	33	97.1	260	6	AX338978	AX338978 Sequence
7	25	73.5	143	1	AY150031	AY150031 Micromono
8	25	73.5	145	1	AY150028	AY150028 Micromono
9	25	73.5	145	1	AY150029	AY150029 Micromono
10	25	73.5	241	6	AX338973	AX338973 Sequence
11	25	73.5	243	6	AX338974	AX338974 Sequence
12	25	73.5	315	6	AX338977	AX338977 Sequence
13	24.6	72.4	34806	1	AF481102	AF481102 Candidatu
14	24	70.6	143	1	AY150030	AY150030 Micromono
15	24	70.6	255	6	AX338976	AX338976 Sequence
16	23.6	69.4	15708	3	AF051097	AF051097 Balanoglo
17	22.4	65.9	180	1	MYCTGH	U39716 Mycoplasma
18	22.4	65.9	10252	1	U39716	U39716 Mycoplasma
19	22.4	65.9	10640	1	AE000029	AE000029 Mycoplasma
20	22.4	65.9	110000	6	AR300198	Continuation (5 of
21	22.4	65.9	155932	1	CJ11168X5	AL139079 Campyloba
22	22.4	65.9	224279	10	AC102135	AC102135 Mus muscu
23	22.2	65.3	215065	2	AC103291	AC103291 Rattus no
24	22.2	65.3	222507	2	AC112440	AC112440 Rattus no
25	22.2	65.3	226214	2	AC126653	AC126653 Rattus no
26	22	64.7	236921	2	AC094151	AC094151 Rattus no
27	22	64.7	241131	2	AC118438	AC118438 Rattus no
28	21.8	64.1	620	8	AF363748	AF363748 Populus d
29	21.8	64.1	174470	10	AC114820	AC114820 Mus muscu
30	21.8	64.1	205196	2	AC131173	AC131173 Rattus no
31	21.8	64.1	222714	2	AC116514	AC116514 Mus muscu
32	21.8	64.1	261157	2	AC103330	AC103330 Rattus no
33	21.8	64.1	284749	2	AC113251	AC113251 Rattus no
34	21.4	62.9	236458	2	AC107428	AC107428 Rattus no
35	21.4	62.9	271197	2	AC117974	AC117974 Rattus no
36	21.2	62.4	110970	10	AC004807	AC004807 Mus muscu
37	21.2	62.4	150986	2	AC023802	AC023802 Mus muscu
38	21.2	62.4	151437	2	AC116850	AC116850 Mus muscu
39	21.2	62.4	181721	2	AC102230	AC102230 Mus muscu
40	21.2	62.4	194695	2	AC026375	AC026375 Mus muscu
41	21.2	62.4	197338	2	AC108265	AC108265 Rattus no
42	21.2	62.4	233242	10	AL591070	AL591070 Mouse DNA
43	21.2	62.4	247759	2	AC106669	AC106669 Rattus no
44	21	61.8	110000	2	AC110649_0	AC110649 Rattus no
45	21	61.8	146984	2	AC144875	AC144875 Pongo pyg

ALIGNMENTS

RESULT 1  
AX338972  
LOCUS AX338972 34 bp DNA linear PAT 09-JAN-2002  
DEFINITION Sequence 3 from Patent WO0187936.  
ACCESSION AX338972  
VERSION AX338972.1 GI:18129108  
KEYWORDS Micromonospora carbonacea  
SOURCE Micromonospora carbonacea  
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Micromonosporaceae; Micromonospora.  
REFERENCE  
1 Hosted, T.J. and Horan, A.C.  
AUTHORS Isolation of Micromonospora carbonacea var africana pm1pl integrase  
TITLE and use of integrating function for site-specific integration into

```

Micromonospora halophitica and Micromonospora carbonacea chromosome
Patent: WO 0187936-A 3 22-NOV-2001;
SCHERING CORPORATION (US)
FEATURES
    source
        1. .34
            /organism="Micromonospora carbonacea"
            /mol_type="unassigned DNA"
            /db_xref="taxon:47853"
ORIGIN
Query Match      100.0%; Score 34; DB 6; Length 34;
Best Local Similarity 100.0%; Pred. No. 0.00015;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCGCGTACGGGTTCAATTCATCCATCAGTCACCG 34
    |||
Db 1 CCGCGTACGGGTTCAATTCATCCATCAGTCACCG 34
    |||
RESULT 2
AV150033      145 bp DNA linear BCT 02-SEP-2003
LOCUS
DEFINITION
Micromonospora sp. ATCC 39149 attr region.
ACCESSION
AV150033
VERSION
AV150033.1 GI:28630429
KEYWORDS
SOURCE
Micromonospora sp. ATCC 39149
ORGANISM
Micromonospora sp. ATCC 39149
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.
REFERENCE
1 (Bases 1 to 145)
AUTHORS
Alexander,D.C., Devlin,D.J., Hewitt,D.D., Horan,A.C. and
Hosted,T.J.
TITLE
Development of the Micromonospora carbonacea var. africana ATCC
39149 bacteriophage pMLP1 integrase for site-specific integration
in Micromonospora spp
JOURNAL
Microbiology (Reading, Engl.) 149 (Pt 9), 2443-2453 (2003)
PUBMED
12949170
REFERENCE
2 (Bases 1 to 145)
AUTHORS
Hosted,T.J. Jr., Hewitt,D.D. and Alexander,D.C.
TITLE
Direct Submission
JOURNAL
Submitted (12-SEP-2002) New Lead Discovery, Schering-Plough
Research, 2015 Galloping Hill Road, Kenilworth, NJ 07033, USA
FEATURES
    source
        56. .145
            /organism="Micromonospora sp. ATCC 39149"
            /mol_type="genomic DNA"
            /strain="ATCC 39149"
            /db_xref="ATCC:39149"
            /db_xref="taxon:219305"
            /focus
            /notes="deposited in ATCC as Micromonospora carbonacea var.
africana Waitz et al."
        1. .55
            /organism="Integration vector pSPRH840"
            /mol_type="genomic DNA"
            /db_xref="taxon:219291"
        1. .145
            /notes="attr region"
        55. .56
            /notes="site of pSPRH840 recombination"
        56. .79
            /notes="pSPRH840 attachment site; attB/attP identity
element"
    stem_loop
        join(85. .98,103. .116)
ORIGIN
Query Match      100.0%; Score 34; DB 1; Length 145;
Best Local Similarity 100.0%; Pred. No. 0.00016;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCGCGTACGGGTTCAATTCATCCATCAGTCACCG 34
    |||

```

```

Db 47 CCGCGTACGGGTTCAATTCATCCATCAGTCACCG 80
RESULT 3
AX338975      247 bp DNA linear PAT 09-JAN-2002
LOCUS
DEFINITION
Sequence 6 from Patent WO0187936.
ACCESSION
AX338975
VERSION
AX338975.1 GI:18129111
KEYWORDS
Micromonospora carbonacea
SOURCE
Micromonospora carbonacea
ORGANISM
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.
REFERENCE
1
AUTHORS
Hosted,T.J. and Horan,A.C.
TITLE
Isolation of Micromonospora carbonacea var africana pmlp1 integrase
and use of integrating function for site-specific integration into
Micromonospora halophitica and Micromonospora carbonacea chromosome
Patent: WO 0187936-A 6 22-NOV-2001;
SCHERING CORPORATION (US)
JOURNAL
Location/Qualifiers
FEATURES
    source
        1. .247
            /organism="Micromonospora carbonacea"
            /mol_type="unassigned DNA"
            /db_xref="taxon:47853"
ORIGIN
Query Match      100.0%; Score 34; DB 6; Length 247;
Best Local Similarity 100.0%; Pred. No. 0.00016;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCGCGTACGGGTTCAATTCATCCATCAGTCACCG 34
    |||
Db 92 CCGCGTACGGGTTCAATTCATCCATCAGTCACCG 125
    |||
RESULT 4
AX150027      2025 bp DNA linear PHG 02-SEP-2003
LOCUS
DEFINITION
Bacteriophage pMLP1 att/int region.
ACCESSION
AX150027
VERSION
AX150027.1 GI:28630421
KEYWORDS
Bacteriophage pMLP1
SOURCE
Bacteriophage pMLP1
ORGANISM
Viruses.
REFERENCE
1 (Bases 1 to 2025)
AUTHORS
Alexander,D.C., Devlin,D.J., Hewitt,D.D., Horan,A.C. and
Hosted,T.J.
TITLE
Development of the Micromonospora carbonacea var. africana ATCC
39149 bacteriophage pMLP1 integrase for site-specific integration
in Micromonospora spp
JOURNAL
Microbiology (Reading, Engl.) 149 (Pt 9), 2443-2453 (2003)
PUBMED
12949170
REFERENCE
2 (Bases 1 to 2025)
AUTHORS
Hosted,T.J. Jr., Alexander,D.C. and Hewitt,D.D.
TITLE
Direct Submission
JOURNAL
Submitted (12-SEP-2002) New Lead Discovery, Schering-Plough
Research, 2015 Galloping Hill Road, Kenilworth, NJ 07033, USA
FEATURES
    source
        1. .2025
            /organism="Bacteriophage pMLP1"
            /mol_type="genomic DNA"
            /specific_host="Micromonospora sp. ATCC 39149"
            /db_xref="taxon:219292"
            /note="host is deposited in ATCC as Micromonospora
carbonacea var. africana Waitz et al."
        1. .2025
            /note="att/int region"
        197. .622
            /note="xis"
        /codon_start=1
    misc_feature
        1. .2025
    CDS
        197. .622

```





Db 56 GGGTTCAATTCCTCAGTCACCG 80

AX338973 241 bp DNA linear PAT 09-JAN-2002

LOCUS AX338973

DEFINITION Sequence 4 from Patent WO0187936.

ACCESSION AX338973

VERSION AX338973.1 GI:18129109

KEYWORDS

SOURCE Micromonospora carbonacea

ORGANISM Bacteria; Actinobacteridae; Actinomycetales; Micromonosporaceae; Micromonospora.

REFERENCE 1

AUTHORS Hosted,T.J. and Horan,A.C.

TITLE Isolation of Micromonospora carbonacea var africana pmpl1 integrase and use of integrating function for site-specific integration into Micromonospora halophytica and Micromonospora carbonacea chromosome Patent: WO 0187936-A 5 22-NOV-2001;

JOURNAL SCHERING CORPORATION (US)

FEATURES

source Location/Qualifiers

1. .241

/organism="Micromonospora carbonacea"

/mol\_type="unassigned DNA"

/db\_xref="taxon:47853"

ORIGIN

Query Match 73.5%; Score 25; DB 6; Length 241;

Best Local Similarity 100.0%; Pred. No. 2.2;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 GGGTTCAATTCCTCAGTCACCG 34

Db 95 GGGTTCAATTCCTCAGTCACCG 119

RESULT 11

AX338974

LOCUS AX338974

DEFINITION Sequence 5 from Patent WO0187936.

ACCESSION AX338974

VERSION AX338974.1 GI:18129110

KEYWORDS

SOURCE Micromonospora carbonacea

ORGANISM Bacteria; Actinobacteridae; Actinomycetales; Micromonosporaceae; Micromonospora.

REFERENCE 1

AUTHORS Hosted,T.J. and Horan,A.C.

TITLE Isolation of Micromonospora carbonacea var africana pmpl1 integrase and use of integrating function for site-specific integration into Micromonospora halophytica and Micromonospora carbonacea chromosome Patent: WO 0187936-A 5 22-NOV-2001;

JOURNAL SCHERING CORPORATION (US)

FEATURES

source Location/Qualifiers

1. .243

/organism="Micromonospora carbonacea"

/mol\_type="unassigned DNA"

/db\_xref="taxon:47853"

ORIGIN

Query Match 73.5%; Score 25; DB 6; Length 243;

Best Local Similarity 100.0%; Pred. No. 2.2;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 GGGTTCAATTCCTCAGTCACCG 34

Db 113 GGGTTCAATTCCTCAGTCACCG 137

RESULT 12

AX338977

LOCUS AX338977

DEFINITION Sequence 8 from Patent WO0187936.

ACCESSION AX338977

VERSION AX338977.1 GI:18129113

KEYWORDS

SOURCE Micromonospora halophytica

ORGANISM Bacteria; Actinobacteridae; Actinomycetales; Micromonosporaceae; Micromonospora.

REFERENCE 1

AUTHORS Hosted,T.J. and Horan,A.C.

TITLE Isolation of Micromonospora carbonacea var africana pmpl1 integrase and use of integrating function for site-specific integration into Micromonospora halophytica and Micromonospora carbonacea chromosome Patent: WO 0187936-A 8 22-NOV-2001;

JOURNAL SCHERING CORPORATION (US)

FEATURES

source Location/Qualifiers

1. .315

/organism="Micromonospora halophytica"

/mol\_type="unassigned DNA"

/db\_xref="taxon:47864"

ORIGIN

Query Match 73.5%; Score 25; DB 6; Length 315;

Best Local Similarity 100.0%; Pred. No. 2.2;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 GGGTTCAATTCCTCAGTCACCG 34

Db 96 GGGTTCAATTCCTCAGTCACCG 120

RESULT 13

AF481102/c

LOCUS AF481102

DEFINITION Candidatus Tremblaya princeps mvin DNA linear BCT 07-JUL-2002

ribosomal protein L13 (rpl13), ribosomal protein S9 (rps9), putative protein HesB (hesB), dihydroxyacid dehydrase (ilvd), 5-enolpyruvylshikimate-3-phosphate synthase (aroA), ribosomal protein S1 (rps1), ribosomal protein S16 (rps16), ribosomal protein L19 (rpl19), 5, 10-methylenetetrahydrofolate reductase (metF), chaperone Hsp60 (groEL), chaperone Hsp10 (groES), DNA polymerase alpha subunit (dnaB), valine sensitive acetolactate synthase III subunit (ilvi), ketol-acid reductoisomerase (ilvc), alpha-isopropylmalate synthase (leuA), and ribosomal protein S15 (rps15) genes, complete cds; 16S ribosomal RNA, 23S ribosomal RNA, and 5S ribosomal RNA genes, complete sequence; putative protein Yabc (yabc), DNA polymerase III epsilon subunit (dnaQ), cell division protein (ftsJ), arginosuccinate lyase (argH), and ribosomal protein L31 (rpl31) genes, complete cds; and unknown genes.

ACCESSION AF481102

VERSION AF481102.1 GI:21702567

KEYWORDS

SOURCE Candidatus Tremblaya princeps

ORGANISM Candidatus Tremblaya Princeps

REFERENCE 1

AUTHORS Baumann,L., Thao,M.L., Hess,J.M., Johnson,M.W. and Baumann,P.

TITLE The genetic properties of the primary endosymbionts of mealybugs differ from those of other endosymbionts of plant sap-sucking insects

JOURNAL Appl. Environ. Microbiol. 68 (7), 3198-3205 (2002)

MEDLINE 22083449

PUBMED 12089995

REFERENCE 2

(bases 1 to 34806)

AUTHORS Baumann,L., Thao,M.L., Hess,J.M. and Baumann,P.

TITLE Direct Submission

JOURNAL Submitted (06-FEB-2002) Microbiology, University of California, One Shields Ave., Davis, CA 95616-8665, USA

FEATURES

source Location/Qualifiers

1. .34806

/organism="Candidatus Tremblaya princeps"

```
/mol_type="genomic DNA"
/db_xref="taxon:189385"
/notes="endosymbiont"
<1..843
/genes="mviN"
<1..843
/genes="mviN"
/notes="putative virulence protein"
/codon_start=1
/transl_table=11
/product="mviN"
/protein_id="AAW75965.1"
/db_xref="GI:21702568"
/translation="EFSTGLYGGSLGSLLPSSRSWASGDCAGYAVLMDWGLSIVL
AGAPNALLFTLLPELTSASFNFGYGRFDHTVQWVAPLAAVPGMGLMGLLRTLSQGL
YARNVSTVPLASLILCSQAAGRLMLPMGHAGLWSSSIGAMLNALLRSLSQ
GAYATPRWASLCARISAAVAGTILSAIARRVDWGMVSURPELRAASVIAAF
AYAFILKGLGLMGMWREASCAASNTVPSNQYKSRAGQARAIALVAVGNRLGVP
FRQA"
1136..1510
/genes="rpl13"
1136..1510
/genes="rpl13"
/codon_start=1
/transl_table=11
/product="ribosomal protein L13"
/protein_id="AAW75966.1"
/db_xref="GI:21702569"
/translation="MKDNGQHVLDARGMVLGRVAVTAVGLLRGKYKASYAHNAAPHNA
VTVLNCHEVRISSKAVGVYVYHSHGPGQLRRTKLHDLMLRSPATVLRAAVHGMLPR
NLRDRPMSMLVHANEGRAR"
1507..1893
/genes="rps9"
1507..1893
/genes="rps9"
/codon_start=1
/transl_table=11
/product="ribosomal protein S9"
/protein_id="AAW75967.1"
/db_xref="GI:21702570"
/translation="MIGWRYGTRRRKRSVARVFLKAAAGTAVVNGMPLGLYFRSAH
AGHAAPLHVTPVAVVAVVQGGIGGAGAVRLGISRALCLSPCLRTIASARLL
TDQRVERKKPKLRARRRQFSKR"
complement (1899..2225)
/genes="hesB"
1899..2225
/codon_start=1
/transl_table=11
/product="putative protein HesB"
/protein_id="AAW75968.1"
/db_xref="GI:21702571"
/translation="MLQINCAVLELRGMVACHGKSCVPLRVFVKGCGKGMHRKFCI
ASRTVGDAKMTLRGLRVADVPLSLMYLRGSIYDCTGVGDARLVVYQPSAGWCSG
ASFTPV"
2317..3981
/genes="ilvD"
2317..3981
/genes="ilvD"
/codon_start=1
/transl_table=11
/product="dihydroxyacid dehydrase"
/protein_id="AAW75969.1"
/db_xref="GI:21702572"
/translation="MNTNGRRITHGFRSPNAMLCAWYDGDFAKPIIGVGCYST
ITPCNAGIORVDAKAMLRHGMATQVFGIPTISDGLSMGTGPMGRYSLVREAVTDC
IFACACGGGLLVVCGCDKXNLPQGMIGLLRANIPGLVYSGTILPGYVWGAELTVV
SFEAVGACGRQMSIDANREVERHACPTGCGMYTANTMSAFEGALGWSLYSST
APSPGEGLSISTARSUDINAVRGIRPDRVYTHRSIRKNAVMVAVGGSNTAVLHY
LAIAAARGALSILHVELIRRVFVPCNNKPSGLSHLADLHSAAGVPRVWHELAGL
IDESCLITGRTIGALLAARHKGSTVTLPTNVALYRTGLVLSGNMNSGAVAK
TSGLSILLHSGTAVFRSEACVQAILNGTCVRIQDVVLIYLPKPGFGPMEPLSPTA
ALVGMGLGQSACLITDGRFSGGTWGLVVGHSPEAAVGSIALVRENGDFITVDLRNS
```

```
CDS
4467..5498
/notes="ORF-A"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAW75970.1"
/db_xref="GI:21702573"
/translation="MOSLATMNLILGAPFVPLTLCSSAWHPICPVPSPNCPHNRKT
AIAVALSFAPNLLARMSRHCNFRFVLYHEQSMGVNDKTMHILIFLAACLRENSG
HPAHRATGKALACACLLQGMHIIROVQLESQPVARCGRKANSTACIACVADGAG
GDMFATLISRDLPAYAKQDGPPIPRKLEAGIEMVAECVPMISNLRDVRRTQOLI
PIVESNILLANRLTPSORARVLMWRCQVDSVHKAAVAPFAGRGFWNAVHIELCWAY
SRYTTELDITTEFYSSYTAGSKYLKRRQVGRVDVRLVKRCTLQHLASLYRIRAE
AIEMLGIIIG"
5499..6500
/notes="ORF-B"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAW75971.1"
/db_xref="GI:21702574"
/translation="MPSFHTELRYINGHDGLSACAARLGVASRIGDMKRAAGMCS
YDARTRETVAREVVVVDVYGGRIAPAPLSLWRCIMNGARSVEPECLALHILGPMCTH
SERATMALRQSORQVQFASFGACSMACRAGCTILPVHNTLSGAVSAAIRVCSVLRL
LAALCSVAVMHPAPSSGRVAVTAOKHALMQCSSWFCDSRMVASVARSASEPSI
AAAQACACATTPALARAMRCHGAALAHSENNRTSFVACSCAGMRSVIGSRVAVMS
LDSRCCECHCAALLIALRHSPSCHVLVIEARAGGAAPVARSRLRVLPQAGCCPQ"
6485..7783
/genes="aroA"
6485..7783
/genes="aroA"
/codon_start=1
/transl_table=11
/product="5-enolpyruvylshikimate-3-phosphate synthase"
/protein_id="AAW75972.1"
/db_xref="GI:21702575"
/translation="MPLPVRSAHVRIHRRRGLALVPTSKSASNRREVACAFVSLKATQ
LSGRTLHSDDTAVLNAHSAFGILABHGRTLRLGPCKGGLLVAELVYVSGSTTAR
LVGASACRRPRHALVHGDDPMRHRPALATVGLASIGAKVLFHFAVQSLVPAIGPTCN
LKNFAPRLPMSESSQFITSASVASHGIYHKDTPACGOLTSRLVYVASTIDIMGLFG
MCESEGQWRYAHAPASMCCEYFGSTDFDFTMSASYPTALCCSGMGMWRLCGVDHASQ
MDADAMASMGCAWQCQCDGNAMWRAARHAPTPRLCCSAPDGMANTAMTISLALAG
ASDAVGLVRYXETDRVLAMALELRMGVAADCNGMSMACUSCLTWHGRADVSTIR
DHRIAMCCSLMHAAGCGVTSRGPQCVCKTYPDYLHECAGT"
7899..9455
/genes="rps1"
7899..9455
/genes="rps1"
/codon_start=1
/transl_table=11
/product="ribosomal protein S1"
/protein_id="AAW75973.1"
/db_xref="GI:21702576"
/translation="MHDHLELRKYDAGPGSIVAAARVVAISSNVHVLRAGLKSD
SRVPLGEFVSPQDRLEVKVGDVTLAYVERDAYDECTVRLRSKRRRVAMWGLDLSAH
RTRGTGTGVGRVKGMTAVVNGMEAFIPGSLIDTRPVKDAERYRGRRLRHEVILGL
RKGNVYLSRVSRIERRQAEENRLLASLGDGVQGVKNITDYGAFIDLGLDGLL
HITDISWRIRHPSOVLVGDRITVKVLKFDTRKTVSLGKQLHLPDPWAGIIGRYPFA
GSRVGRVTNIADYGLFVEIERGIEGLLHVSEMHKSGKTSILRHFRVQYVNVTVLD
VSKTRRIISLGMKHGTLNPLWTFYKFRKNDRMGAIIRSPDVGLVDLPMGVHGVVC
MSKDLRPLRALCSGSGVVEARVVGVDVNRGRISLASGALGPLGVRIILAPGFIIRPGA
FVMAAPMRECCILNVRAPCAAAKQAQALRGAVHARLLLSAISIRPSRSVACAIRGMR
ANHGLVPCLP"
9278..9724
/genes="rps16"
9278..9724
/genes="rps16"
/codon_start=1
/transl_table=11
/product="ribosomal protein S16"
/protein_id="AAW75974.1"
/db_xref="GI:21702577
```

```

/translation="MLPQASSAVRCQCPVRSRSDRCDAKCYFTTSLRCMRHS
RYRASKPWACSLAIRRCGARGRPAYQVVVADSRKRRNGVFLCRGYNPRLUKSAH
IDTVMLRWTERGAAPTETVAELLHRHAAGCTAAARNSGSVCPHMQ"
9706..110020
/gene="rpl19"
9706..110020
/gene="rpl19"
/codon_start=1
/transl_table=11
/product="ribosomal protein L19"
/protein_id="RAAM75975.1"
/db_xref="GI:21702578"
/translacion="MSTAIARVLEARTMGAATDAPLPYPRIGDVAISGTTGRGKGY
APEGVVSRRSALNTSIVVREHAGFCVDLALRLHSHDVPARSARSTSCNAPACVC
RP"

Query Match      72.4%; Score 24.6; DB 1; Length 34806;
Best Local Similarity 87.1%; Pred. No. 3.6;
Matches 27; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 CCGGTACGGTTCATTCCTCCATCAGTCACCC 33
DB 32511 CAGGTATGGTTCGAGTCCCATCAGTCACCC 32481

RESULT 14
AY150030      143 bp      DNA      linear      BCT 02-SEP-2003
LOCUS      Micromonospora nigra attB region.
DEFINITION      Micromonospora nigra attB region.
ACCESSION      AY150030
VERSION      AY150030.1 GI:28630426
KEYWORDS
SOURCE      Micromonospora nigra
ORGANISM      Micromonospora nigra
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.
REFERENCE      1 (bases 1 to 143)
AUTHORS      Alexander,D.C., Devlin,D.J., Hewitt,D.D., Horan,A.C. and
Hosted,T.J.
TITLE      Development of the Micromonospora carbonacea var. africana ATCC
39149 bacteriophage pMPL1 integrase for site-specific integration
in Micromonospora spp
JOURNAL      Microbiology (Reading, Engl.) 149 (Pt 9), 2443-2453 (2003)
PUBMED      12949170
REFERENCE      2 (bases 1 to 143)
AUTHORS      Hosted,T.J. Jr., Hewitt,D.D. and Alexander,D.C.
TITLE      Direct Submission
JOURNAL      Submitted (12-SEP-2002) New Lead Discovery, Schering-Plough
Research, 2015 Galloping Hill Road, Kenilworth, NJ 07033, USA
FEATURES
source      Location/Qualifiers
1..143
/organism="Micromonospora nigra"
/mol_type="genomic DNA"
/strain="ATCC 33088"
/db_xref="ATCC:33088"
/db_xref="taxon:145857"
misc_feature      1..143
trna      5..79
/product="tRNA-His"
misc_feature      56..80
stem_loop      /note="attachment site; attB/attP identity element"
join(94..105,110..121)
ORIGIN

Query Match      70.6%; Score 24; DB 1; Length 143;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GGGTTCAATTCCTCCATCAGTCACCC 33
DB 56 GGGTTCAATTCCTCCATCAGTCACCC 79

/translation="MLPQASSAVRCQCPVRSRSDRCDAKCYFTTSLRCMRHS
RYRASKPWACSLAIRRCGARGRPAYQVVVADSRKRRNGVFLCRGYNPRLUKSAH
IDTVMLRWTERGAAPTETVAELLHRHAAGCTAAARNSGSVCPHMQ"
9706..110020
/gene="rpl19"
9706..110020
/gene="rpl19"
/codon_start=1
/transl_table=11
/product="ribosomal protein L19"
/protein_id="RAAM75975.1"
/db_xref="GI:21702578"
/translacion="MSTAIARVLEARTMGAATDAPLPYPRIGDVAISGTTGRGKGY
APEGVVSRRSALNTSIVVREHAGFCVDLALRLHSHDVPARSARSTSCNAPACVC
RP"

Query Match      72.4%; Score 24.6; DB 1; Length 34806;
Best Local Similarity 87.1%; Pred. No. 3.6;
Matches 27; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 CCGGTACGGTTCATTCCTCCATCAGTCACCC 33
DB 32511 CAGGTATGGTTCGAGTCCCATCAGTCACCC 32481

RESULT 15
AX338976      255 bp      DNA      linear      PAT 09-JAN-2002
LOCUS      Sequence 7 from Patent WO0187936.
DEFINITION      AX338976
ACCESSION      AX338976
VERSION      AX338976.1 GI:18129112
KEYWORDS      Micromonospora halophytica
SOURCE      Micromonospora halophytica
ORGANISM      Micromonospora halophytica
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.
REFERENCE      1
AUTHORS      Hosted,T.J. and Horan,A.C.
TITLE      Isolation of Micromonospora carbonacea var africana pmlp1 integrase
and use of integrating function for site-specific integration into
Micromonospora halophytica and Micromonospora carbonacea chromosome
Patent: WO 0187936-A 7 22-NOV-2001;
SCHERING CORPORATION (US)
FEATURES
source      Location/Qualifiers
1..255
/organism="Micromonospora halophytica"
/mol_type="unassigned DNA"
/db_xref="taxon:47864"
ORIGIN

Query Match      70.6%; Score 24; DB 6; Length 255;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GGGTTCAATTCCTCCATCAGTCACCC 33
DB 96 GGGTTCAATTCCTCCATCAGTCACCC 119

Search completed: August 6, 2004, 11:53:13
Job time : 159.122 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2004, 08:18:49 ; Search time 20.205 Seconds  
(without alignments)  
7148.662 Million cell updates/sec

Title: US-09-855-340A-3  
Perfect score: 34  
Sequence: 1 cccggtagcgggtcaattcccatcagtcacccg 34

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_25Jan04: \*  
1: Geneseq1980s: \*  
2: Geneseq1990s: \*  
3: Geneseq2000s: \*  
4: Geneseq2001as: \*  
5: Geneseq2001bs: \*  
6: Geneseq2001bs: \*  
7: Geneseq2002s: \*  
8: Geneseq2003as: \*  
9: Geneseq2003bs: \*  
10: Geneseq2004s: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	100.0	34	6	Aad25933 Micromono
2	34	100.0	209	6	Aad25945 Micromono
3	34	100.0	247	6	Aad25936 Micromono
4	34	100.0	4388	5	Aas08694 Micromono
5	33	97.1	260	6	Aad25939 Micromono
6	25	73.5	241	6	Aad25934 Micromono
7	25	73.5	243	6	Aad25935 Micromono
8	25	73.5	315	6	Aad25938 Micromono
9	24	70.6	255	6	Aad25937 Micromono
10	22.4	65.9	110000	2	Continuation (5 of AAT58840_4
C 11	20.6	60.6	5546	4	AB112144 Drosophil
C 12	20.4	60.0	40116	7	AB226080 Mouse Dna
13	19.6	57.6	348	5	Aaf15421 Thalassio
14	19.6	57.6	348	5	Aaf15526 Thalassio
C 15	19.6	57.6	438	8	Ach26679 Human adu
C 16	19.6	57.6	556	8	Ach28010 Human adu
C 17	19.6	57.6	40681	6	Continuation (7 of ABA92787_6
C 18	19.6	57.6	171936	6	ABes5655 Human SUL
C 19	19.4	57.1	532	3	Aae62850 Human pro
C 20	19.4	57.1	6573	5	Aah4107 Human per
C 21	19.4	57.1	9883	4	Aak79597 Human inn
C 22	19.4	57.1	17601	6	AB152077 Human per
C 23	19.4	57.1	50196	4	Aak79598 Human inn

## ALIGNMENTS

## RESULT 1

AAD25933  
ID AAD25933 standard; DNA; 34 BP.

XX AC AAD25933;

XX AC AAD25933;

DT 26-MAR-2002 (first entry)

XX DE Micromonospora carbonacea pMLP1 attP site DNA.

XX KW Integrase; int; excisionase; xis; integrase attachment site; attP; pMLP1;

XX KW site-specific integration; hybrid antibiotic; metabolic product;

XX KW secondary metabolic pathway; ds.

XX OS Micromonospora carbonacea.

XX PN WO200187936-A2.

XX PD 22-NOV-2001.

XX PF 15-MAY-2001; 2001WO-US015760.

XX PR 17-MAY-2000; 2000US-0204670P.

XX PA (SCHE ) SCHERING CORP.

XX PI Hosted TJ, Horan AC;

XX DR WPI; 2002-082983/11.

XX PT Novel polynucleotides encoding integrase, excisionase and an integrase

XX PT attachment site isolated from a lysogenic phase pMLP1, useful for

XX PS Claim 5; Page 33; 34pp; English.

XX CC The present invention relates to novel polynucleotides encoding integrase

XX CC (int) and excisionase (xis) and an integrase attachment site (attP) which

XX CC are isolated from pMLP1, a bacteriophage (lysogenic phase) isolated from

XX CC Micromonospora carbonacea var. africana. Polynucleotides of the invention

XX CC are useful for transforming an actinomycete with a vector. They are also

XX CC useful for creating vectors for site-specific integration into host

XX CC chromosomes. The integrating vectors are used to express actinomycete

XX CC genes, manipulate secondary metabolic pathways and create new metabolic

XX CC products such as hybrid antibiotics. The present sequence is pMLP1 attP

XX CC site DNA from Micromonospora carbonacea

Abk81114 Bacillus  
Abk81134 Bacillus  
Abk79564 Bacillus  
Aah07734 Human CDN  
Aai95794 Human neu  
Aah84379 E. Coli g  
Aax13475 Enterococ  
Abs99270 Enterococ  
Aah14935 Human CDN  
Abl90292 Human pol  
Aas86629 DNA encod  
Aax20518 Polynucle  
Ada02693 Mouse Tk2  
Adb72431 Mouse Tk2  
Ada02627 Mouse Flt  
Adb72365 Mouse Flt  
Aaq04525 Total bas  
Aaq04525 Total bas  
Aab23503 Cellobioh  
Aaz65283 Human sec  
Ada40018 Human sec  
Adb91229 Human sec

24 19.2 56.5 76 6 ABK81114  
25 19.2 56.5 76 6 ABK81134  
26 19.2 56.5 171 6 ABK79564  
27 19.2 56.5 744 4 AAH07734  
28 19.2 56.5 744 4 AAI95794  
29 19.2 56.5 839 4 AAH84379  
30 19.2 56.5 987 2 AAX13475  
31 19.2 56.5 987 6 ABS99270  
32 19.2 56.5 1433 4 AAH14935  
33 19.2 56.5 1723 6 ABL90292  
34 19.2 56.5 4443 5 AAS86629  
35 19.2 56.5 19217 2 AAX20518  
36 19.2 56.5 41936 8 ADA02693  
37 19.2 56.5 41936 9 ADB72431  
38 19.2 56.5 47115 8 ADA02627  
39 19.2 56.5 47115 9 ADB72365  
40 19.2 56.5 134525 2 AAQ04525  
41 19.2 56.5 134525 2 AAQ04525  
42 19 55.9 1581 7 ABT23503  
43 19 55.9 1680 3 AAZ65283  
44 19 55.9 1681 7 ADA40018  
45 19 55.9 1681 8 ADB91229

```

XX SQ Sequence 34 BP; 6 A; 14 C; 7 G; 7 T; 0 U; 0 Other;
Query Match 100.0%; Score 34; DB 6; Length 34;
Best Local Similarity 100.0%; Pred. No. 1.8e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCCGGTACGGTTCAATTCCCATCAGTCACCG 34
D5 1 CCCCGGTACGGTTCAATTCCCATCAGTCACCG 34

RESULT 2
AAD25945
ID AAD25945 standard; DNA; 209 BP.
XX AC AAD25945;
XX DT 26-MAR-2002 (first entry)
XX DE Micromonospora carbonacea pMLP1 attP DNA.
XX KW Integrase; int; excisionase; xis; integrase attachment site; attP; pMLP1;
XX KW site-specific integration; hybrid antibiotic; metabolic product;
XX KW secondary metabolic pathway; ds.
XX OS Micromonospora carbonacea.
XX FH Key Location/Qualifiers
XX FT misc_feature 101..125
XX FT /*tag= a
XX FT /*note= "attP region"
XX FT repeat_region 145..162
XX FT /*tag= b
XX FT /rpt_type= INVERTED
XX FT /*note= "Inverted repeat 1 (IR1)"
XX FT repeat_region 167..186
XX FT /*tag= c
XX FT /rpt_type= INVERTED
XX FT /*note= "Inverted repeat 2 (IR2)"
XX WO200187936-A2.
XX 22-NOV-2001.
XX 15-MAY-2001; 2001WO-US015760.
XX 17-MAY-2000; 2000US-0204670P.
XX (SCHE ) SCHERING CORP.
XX Hosted TJ, Horan AC;
XX WPI; 2002-082983/11.
XX Novel polynucleotides encoding integrase, excisionase and an integrase
XX attachment site isolated from a lysogenic phage pMLP1, useful for
XX transforming an actinomycete.
XX Example 3; Fig 4; 34pp; English.
XX The present invention relates to novel polynucleotides encoding integrase
XX (int) and excisionase (xis) and an integrase attachment site (attP) which
XX are isolated from pMLP1, a bacteriophage (lysogenic phage) isolated from
XX Micromonospora carbonacea var. africana. Polynucleotides of the invention
XX are useful for transforming an actinomycete with a vector. They are also
XX useful for creating vectors for site-specific integration into host
XX chromosomes. The integrating vectors are used to express actinomycete
XX genes, manipulate secondary metabolic pathways and create new metabolic
XX products such as hybrid antibiotics. The present sequence is pMLP1 attP
XX DNA from Micromonospora carbonacea
XX Sequence 209 BP; 49 A; 64 C; 66 G; 30 T; 0 U; 0 Other;

Query Match 100.0%; Score 34; DB 6; Length 209;
Best Local Similarity 100.0%; Pred. No. 2.5e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCCGGTACGGTTCAATTCCCATCAGTCACCG 34
D5 92 CCCCGGTACGGTTCAATTCCCATCAGTCACCG 125

RESULT 3
AAD25936
ID AAD25936 standard; DNA; 247 BP.
XX AC AAD25936;
XX DT 26-MAR-2002 (first entry)
XX DE Micromonospora carbonacea pMLP1 attP/attB right juncture DNA.
XX KW Integrase; int; excisionase; xis; integrase attachment site; attP; pMLP1;
XX KW site-specific integration; hybrid antibiotic; metabolic product; attB;
XX KW secondary metabolic pathway; attB/attP DNA; ds.
XX OS Micromonospora carbonacea.
XX FH Key Location/Qualifiers
XX FT misc_feature 101..125
XX FT /*tag= a
XX FT /*note= "attP/attB region"
XX FT repeat_region 130..143
XX FT /*tag= b
XX FT /rpt_type= INVERTED
XX FT /*note= "Inverted repeat 1 (IR1)"
XX FT repeat_region 148..161
XX FT /*tag= c
XX FT /rpt_type= INVERTED
XX FT /*note= "Inverted repeat 2 (IR2)"
XX FT CDS complement(185..247)
XX FT /*tag= d
XX FT /product= "attB peptide"
XX WO200187936-A2.
XX 22-NOV-2001.
XX 15-MAY-2001; 2001WO-US015760.
XX 17-MAY-2000; 2000US-0204670P.
XX (SCHE ) SCHERING CORP.
XX Hosted TJ, Horan AC;
XX WPI; 2002-082983/11.
XX P-PSDB; AAE15909.
XX Novel polynucleotides encoding integrase, excisionase and an integrase
XX attachment site isolated from a lysogenic phage pMLP1, useful for
XX transforming an actinomycete.
XX Claim 23; Fig 4; 34pp; English.
XX The present invention relates to novel polynucleotides encoding integrase
XX (int) and excisionase (xis) and an integrase attachment site (attP) which
XX are isolated from pMLP1, a bacteriophage (lysogenic phage) isolated from
XX Micromonospora carbonacea var. africana. Polynucleotides of the invention
XX are useful for transforming an actinomycete with a vector. They are also
XX useful for creating vectors for site-specific integration into host
XX chromosomes. The integrating vectors are used to express actinomycete
XX genes, manipulate secondary metabolic pathways and create new metabolic
XX products such as hybrid antibiotics. The present sequence is pMLP1
XX attP/attB right juncture DNA from Micromonospora carbonacea

```

CC from modifications of the DNA sequence designed to change glycosyl and  
CC modified orsellinic acid groups contained in everninomicin, for  
CC expressing functional or mutant everninomicin biosynthetic enzyme for  
CC evaluation, diagnosis and preferably biosynthesis of everninomicin or  
CC other secondary metabolic products, improving the yield of everninomicins  
CC and to produce novel everninomicins and also as a hybridisation probe to  
CC identify homologous sequences. The encoded polypeptides are useful for  
CC combinatorial biosynthesis to generate libraries of orthomycins, e.g.  
CC everninomicin analogues/homologues and drug discovery. The DNA encoding  
CC the integrase allows for increasing a given gene dosage. The integrative  
CC vector can be used to permanently integrate copies of a heterologous gene  
CC of choice into chromosomes of different hosts and to integrate genes  
CC which increase the yield of known products or to generate novel products  
CC such as hybrid antibiotics or other novel secondary metabolites. The  
CC vector can also be used to integrate antibiotic resistance genes in order  
CC to carry out bioconversions with compounds to which the strain is  
CC normally sensitive and is thus useful in fermentation processes involving  
CC e.g. Streptomyces antibioticus. (Updated on 11-SEP-2003 to standardise OS  
CC field)  
XX  
SQ Sequence 4388 BP; 714 A; 1468 C; 1506 G; 594 T; 0 U; 6 Other;  
Query Match 100.0%; Score 34; DB 5; Length 4388;  
Best Local Similarity 100.0%; Pred. No. 4.3e-05;  
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CCCCGGTACGGTTCATTCCTCCATCAGTCACCCG 34  
DB 2682 CCCCGGTACGGTTCATTCCTCCATCAGTCACCCG 2715  
RESULT 5  
AAD25939  
ID AAD25939 standard; DNA; 260 BP.  
XX AC AAD25939;  
XX DT 07-AUG-2003 (revised)  
XX DT 26-MAR-2002 (first entry)  
XX DE Micromonospora halophytica pMLP1 attP/attB right juncture DNA.  
XX KW Integrase; int; excisionase; xis; integrase attachment site; attP; pMLP1;  
XX KW site-specific integration; hybrid antibiotic; metabolic product; attB;  
XX KW secondary metabolic pathway; attB/attP DNA; ds.  
XX OS Micromonospora halophytica.  
XX FH Key Location/Qualifiers  
FT misc\_feature 101..126  
FT /tag= a  
FT /note= "attP/attB region"  
FT repeat\_region 139..150  
FT /tag= b  
FT /rpt\_type= INVERTED  
FT repeat\_region 155..166  
FT /note= "Inverted repeat 1 (IR1)"  
FT /tag= c  
FT /rpt\_type= INVERTED  
FT /note= "Inverted repeat 2 (IR2)"  
FT CDS complement(196..258)  
FT /tag= d  
FT /product= "attB peptide".  
XX WO200187935-A2.  
XX PN 22-NOV-2001.  
XX PD 15-MAY-2001; 2001WO-US015760.  
XX PF 17-MAY-2000; 2000US-0204670P.  
XX PR (SCHE ) SCHERING CORP.  
XX PA

XX SQ Sequence 247 BP; 38 A; 82 C; 90 G; 37 T; 0 U; 0 Other;  
Query Match 100.0%; Score 34; DB 6; Length 247;  
Best Local Similarity 100.0%; Pred. No. 2.6e-05;  
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CCCCGGTACGGTTCATTCCTCCATCAGTCACCCG 34  
DB 92 CCCCGGTACGGTTCATTCCTCCATCAGTCACCCG 125  
RESULT 4  
AAS08694  
ID AAS08694 standard; DNA; 4388 BP.  
XX AC AAS08694;  
XX DT 11-SEP-2003 (revised)  
XX DT 26-SEP-2001 (first entry)  
XX DE Micromonospora DNA encoding integrase enzymes.  
XX KW Everninomicin; antibiotic; bottle-neck gene; orthomycin; fermentation;  
XX KW integrase; ds.  
XX OS Micromonospora sp. ATCC 39149.  
XX FH Key Location/Qualifiers  
FT CDS 963..1387  
FT /tag= a  
FT /product= "Integrase #1"  
FT CDS 1394..2572  
FT /tag= b  
FT /product= "Integrase #2"  
FT misc\_feature 2570..2799  
FT /tag= c  
FT /note= "AttB/AttP region of integrase action"  
FT misc\_recomb 27114..2715  
FT /tag= d  
FT /label= Insertion\_juncture  
FT /note= "Site of integrase activity"  
XX WO200151639-A2.  
XX PN 19-JUL-2001.  
XX PD 12-JAN-2001; 2001WO-US001187.  
XX PF 12-JAN-2000; 2000US-0175751P.  
XX PR (SCHE ) SCHERING CORP.  
XX PI Hosted TU, Horan AC, Wang TX;  
XX WPI; 2001-442147/47.  
XX P-PSDB; AAU04900, AAU04912.  
XX New nucleic acid molecules encoding everninomicin pathway gene products,  
XX useful for improving yields of everninomicin, to produce new  
XX everninomicin and as probes to identify homologous sequences.  
XX Claim 26; Fig 7; 109pp; English.  
XX The sequence encodes 2 integrases which permit site specific integration  
XX of a vector into an actinomycete, especially a Micromonospora genome.  
XX The invention relates to nucleic acids and vectors comprising a M.  
XX carbonacea everninomicin biosynthetic pathway resistance gene product  
XX useful for selecting for a transfected or transformed host cell. An  
XX integrative version of the vector is useful for introducing a  
XX everninomicin pathway gene (a bottle-neck gene) into an actinomycete of  
XX the genus Micromonospora. The DNA encoding the biosynthetic proteins is  
XX useful for synthesising novel everninomicin-related compounds, arising

```

XX PI Hosted TJ, Horan AC;
XX DR WPI; 2002-082983/11.
XX DR P-PSDB; AAE15910.
XX PT Novel polynucleotides encoding integrase, excisionase and an integrase
XX PT attachment site isolated from a lysogenic phage pMLP1, useful for
XX PT transforming an actinomycete.
XX PS Claim 23; Fig 5; 34pp; English.
XX CC The present invention relates to novel polynucleotides encoding integrase
XX CC (int) and excisionase (xis) and an integrase attachment site (attP) which
XX CC are isolated from pMLP1, a bacteriophage (lysogenic phage) isolated from
XX CC Micromonospora carbonacea var. africana. Polynucleotides of the invention
XX CC are useful for transforming an actinomycete with a vector. They are also
XX CC useful for creating vectors for site-specific integration into host
XX CC chromosomes. The integrating vectors are used to express actinomycete
XX CC genes, manipulate secondary metabolic pathways and create new metabolic
XX CC products such as hybrid antibiotics. The present sequence is pMLP1
XX CC attP/attB right junction DNA from Micromonospora halophytica. (Updated on
XX CC 07-AUG-2003 to correct OS field.)
XX SQ Sequence 260 BP; 47 A; 83 C; 90 G; 40 T; 0 U; 0 Other;
Query Match 97.1%; Score 33; DB 6; Length 260;
Best Local Similarity 100.0%; Pred. No. 7.2e-05;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CCCCGTACGGTTCATTCCTCCATCAGTCACCC 33
Db 92 CCCCGTACGGTTCATTCCTCCATCAGTCACCC 124
RESULT 6
AAD25934
ID AAD25934 standard; DNA; 241 BP.
XX AC AAD25934;
XX DT 26-MAR-2002 (first entry)
XX DE Micromonospora carbonacea pMLP1 attB DNA.
XX KW Integrase; int; excisionase; xis; integrase attachment site; attP; pMLP1;
XX KW site-specific integration; hybrid antibiotic; metabolic product; attB;
XX KW secondary metabolic pathway; ds.
XX OS Micromonospora carbonacea.
XX FH Key Location/Qualifiers
XX FT misc_feature 44..119
XX FT /*tag= a
XX FT /note= "tRNA-His DNA"
XX FT /*tag= b
XX FT /note= "attB region"
XX FT /*tag= c
XX FT /rpt_type= INVERTED
XX FT /note= "Inverted repeat 1 (IR1)"
XX FT /*tag= d
XX FT /rpt_type= INVERTED
XX FT /note= "Inverted repeat 2 (IR2)"
XX FT complement(179..241)
XX FT /*tag= e
XX FT /product= "attB peptide"
XX PN WO200187936-A2.
XX PD 22-NOV-2001.

```

```

XX PF 15-MAY-2001; 2001WO-US015760.
XX PR 17-MAY-2000; 2000US-0204670P.
XX PA (SCHE ) SCHERING CORP.
XX PI Hosted TJ, Horan AC;
XX DR WPI; 2002-082983/11.
XX DR P-PSDB; AAE15909.
XX CC Novel polynucleotides encoding integrase, excisionase and an integrase
XX CC attachment site isolated from a lysogenic phage pMLP1, useful for
XX CC transforming an actinomycete.
XX PS Claim 23; Fig 4; 34pp; English.
XX CC The present invention relates to novel polynucleotides encoding integrase
XX CC (int) and excisionase (xis) and an integrase attachment site (attP) which
XX CC are isolated from pMLP1, a bacteriophage (lysogenic phage) isolated from
XX CC Micromonospora carbonacea var. africana. Polynucleotides of the invention
XX CC are useful for transforming an actinomycete with a vector. They are also
XX CC useful for creating vectors for site-specific integration into host
XX CC chromosomes. The integrating vectors are used to express actinomycete
XX CC genes, manipulate secondary metabolic pathways and create new metabolic
XX CC products such as hybrid antibiotics. The present sequence is pMLP1 attB
XX CC DNA from Micromonospora carbonacea
XX SQ Sequence 241 BP; 27 A; 78 C; 92 G; 44 T; 0 U; 0 Other;
Query Match 73.5%; Score 25; DB 6; Length 241;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 10 GGGTTCAATTCCTCCATCAGTCACCCG 34
Db 95 GGGTTCAATTCCTCCATCAGTCACCCG 119
RESULT 7
AAD25935
ID AAD25935 standard; DNA; 243 BP.
XX AC AAD25935;
XX DT 26-MAR-2002 (first entry)
XX DE Micromonospora carbonacea pMLP1 attB/attP left junction DNA.
XX KW Integrase; int; excisionase; xis; integrase attachment site; attP; pMLP1;
XX KW site-specific integration; hybrid antibiotic; metabolic product; attB;
XX KW secondary metabolic pathway; attB/attP DNA; ds.
XX OS Micromonospora carbonacea.
XX FH Key Location/Qualifiers
XX FT misc_feature 62..137
XX FT /*tag= a
XX FT /note= "tRNA-His DNA"
XX FT /*tag= b
XX FT /note= "attB/attP region"
XX FT /*tag= c
XX FT /rpt_type= INVERTED
XX FT /note= "Inverted repeat 1 (IR1)"
XX FT /*tag= d
XX FT /rpt_type= INVERTED
XX FT /note= "Inverted repeat 2 (IR2)"
XX PN WO200187936-A2.

```

XX 22-NOV-2001.  
 XX 15-MAY-2001; 2001WO-US015760.  
 XX 17-MAY-2000; 2000US-0204670P.  
 XX (SCHE ) SCHERING CORP.  
 XX Hosted TJ, Horan AC;  
 XX WPI; 2002-082983/11.  
 XX Novel polynucleotides encoding integrase, excisionase and an integrase  
 XX attachment site isolated from a lysogenic phage pMLP1, useful for  
 XX transforming an actinomycete.  
 XX Claim 23; Fig 4; 34pp; English.  
 XX The present invention relates to novel polynucleotides encoding integrase  
 XX (int) and excisionase (xis) and an integrase attachment site (attP) which  
 XX are isolated from pMLP1, a bacteriophage (lysogenic phage) isolated from  
 XX Micromonospora carbonacea var. africana. Polynucleotides of the invention  
 XX are useful for transforming an actinomycete with a vector. They are also  
 XX useful for creating vectors for site-specific integration into host  
 XX chromosomes. The integrating vectors are used to express actinomycete  
 XX genes, manipulate secondary metabolic pathways and create new metabolic  
 XX products such as hybrid antibiotics. The present sequence is pMLP1  
 XX attB/attP left junction DNA from Micromonospora carbonacea  
 XX Sequence 243 BP; 39 A; 74 C; 83 G; 47 T; 0 U; 0 Other;  
 SQ Query Match 73.5%; Score 25; DB 6; Length 243;  
 Best Local Similarity 100.0%; Pred.No.0.27; Mismatches 0; Indels 0; Gaps 0;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 10 GGGTTCAATTCCCATCAGTCACCCG 34  
 Db 113 GGGTTCAATTCCCATCAGTCACCCG 137  
 RESULT 8  
 AAD25938  
 ID AAD25938 standard; DNA; 315 BP.  
 AC AAD25938;  
 XX 07-AUG-2003 (revised)  
 DT 26-MAR-2002 (first entry)  
 XX Micromonospora halophytica pMLP1 attB/attP left junction DNA.  
 DE Integrage; int; excisionase; xis; integrase attachment site; attP; pMLP1;  
 KW site-specific integration; hybrid antibiotic; metabolic product; attB;  
 KW secondary metabolic pathway; attB/attP DNA; ds.  
 XX Micromonospora halophytica.  
 OS Key Location/Qualifiers  
 FH misc\_feature 45..120  
 FT /tag= a  
 FT /note= "CRNA-His DNA"  
 FT misc\_feature 96..120  
 FT /tag= b  
 FT /note= "attB/attP region"  
 FT repeat\_region 140..158  
 FT /tag= c  
 FT /rpt\_type= INVERTED  
 FT repeat\_region 162..178  
 FT /note= "Inverted repeat 1 (IR1)"  
 FT /tag= d  
 FT /rpt\_type= INVERTED  
 FT repeat\_region 162..178  
 FT /note= "Inverted repeat 2 (IR2)"

XX WO200187936-A2.  
 XX 22-NOV-2001.  
 XX 15-MAY-2001; 2001WO-US015760.  
 XX 17-MAY-2000; 2000US-0204670P.  
 XX (SCHE ) SCHERING CORP.  
 XX Hosted TJ, Horan AC;  
 XX WPI; 2002-082983/11.  
 XX Novel polynucleotides encoding integrase, excisionase and an integrase  
 XX attachment site isolated from a lysogenic phage pMLP1, useful for  
 XX transforming an actinomycete.  
 XX Claim 23; Fig 5; 34pp; English.  
 XX The present invention relates to novel polynucleotides encoding integrase  
 XX (int) and excisionase (xis) and an integrase attachment site (attP) which  
 XX are isolated from pMLP1, a bacteriophage (lysogenic phage) isolated from  
 XX Micromonospora carbonacea var. africana. Polynucleotides of the invention  
 XX are useful for transforming an actinomycete with a vector. They are also  
 XX useful for creating vectors for site-specific integration into host  
 XX chromosomes. The integrating vectors are used to express actinomycete  
 XX genes, manipulate secondary metabolic pathways and create new metabolic  
 XX products such as hybrid antibiotics. The present sequence is pMLP1  
 XX attB/attP left junction DNA from Micromonospora halophytica. (Updated on  
 XX 07-AUG-2003 to correct OS field.)  
 SQ Sequence 315 BP; 42 A; 109 C; 101 G; 63 T; 0 U; 0 Other;  
 Query Match 73.5%; Score 25; DB 6; Length 315;  
 Best Local Similarity 100.0%; Pred.No.0.29; Mismatches 0; Indels 0; Gaps 0;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 10 GGGTTCAATTCCCATCAGTCACCCG 34  
 Db 96 GGGTTCAATTCCCATCAGTCACCCG 120  
 RESULT 9  
 AAD25937  
 ID AAD25937 standard; DNA; 255 BP.  
 AC AAD25937;  
 XX 07-AUG-2003 (revised)  
 DT 26-MAR-2002 (first entry)  
 XX Micromonospora halophytica pMLP1 attB DNA.  
 DE Integrage; int; excisionase; xis; integrase attachment site; attP; pMLP1;  
 KW site-specific integration; hybrid antibiotic; metabolic product; attB;  
 KW secondary metabolic pathway; ds.  
 XX Micromonospora halophytica.  
 OS Key Location/Qualifiers  
 FH misc\_feature 45..121  
 FT /tag= a  
 FT /note= "tRNA-His DNA"  
 FT misc\_feature 96..121  
 FT /tag= b  
 FT /note= "attB region"  
 FT repeat\_region 134..145  
 FT /tag= c  
 FT /rpt\_type= INVERTED  
 FT repeat\_region 150..161  
 FT /note= "Inverted repeat 1 (IR1)"

FT FT /\*tag= d  
FT /rpt\_type= INVERTED  
FT /note= "Inverted repeat 2 (IR2)"  
CDS complement(191..253)  
FT /\*tag= e  
FT /product= "attB peptide"  
FT

XX WO200187936-A2.

XX 22-NOV-2001.

XX 15-MAY-2001; 2001WO-US015760.

XX 17-MAY-2000; 2000US-0204670P.

XX (SCHE ) SCHERING CORP.

XX Hosted TJ, Horan AC;

XX WPI; 2002-082983/11.

XX P-PSDB; AAE15910.

XX Novel polynucleotides encoding integrase, excisionase and an integrase  
PT attachment site isolated from a lysogenic phage pMLP1, useful for  
PT transforming an actinomycete.

XX Claim 23; Fig 5; 34pp; English.

XX The present invention relates to novel polynucleotides encoding integrase  
CC (int) and excisionase (xis) and an integrase attachment site (attP) which  
CC are isolated from pMLP1, a bacteriophage (lysogenic phage) isolated from  
CC Micromonospora carbonacea var. africana. Polynucleotides of the invention  
CC are useful for transforming an actinomycete with a vector. They are also  
CC useful for creating vectors for site-specific integration into host  
CC chromosomes. The integrating vectors are used to express actinomycete  
CC genes, manipulate secondary metabolic pathways and create new metabolic  
CC products such as hybrid antibiotics. The present sequence is pMLP1 attB  
CC DNA from Micromonospora halophytica. (Updated on 07-AUG-2003 to correct  
CC OS field.)

XX Sequence 255 BP; 33 A; 78 C; 94 G; 50 T; 0 U; 0 Other;

Query Match 70.8%; Score 24; DB 6; Length 255;

Best Local Similarity 100.0%; Pred. No. 0.78;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GGGTTCATTCCCATCAGTCACCC 33

Db 96 GGGTTCATTCCCATCAGTCACCC 119

RESULT 10

Continuation (5 of 6) of AAT58840 from base 400001 (Mycoplasma genitalium genome.)  
WP Sequence split into 6 fragments LOCUS AAT58840 Accession AAT58840

Fragment Name	Begin	End
WP AAT58840_0	1	110000
WP AAT58840_1	100001	210000
WP AAT58840_2	200001	310000
WP AAT58840_3	300001	410000
WP AAT58840_4	400001	510000
WP AAT58840_5	500001	580073

Query Match 65.8%; Score 22.4; DB 2; Length 110000;

Best Local Similarity 95.8%; Pred. No. 12;

Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 10 GGGTTCATTCCCATCAGTCACCC 33

Db 45128 GGGTTCATTCCCATCAGTCACCC 45151

RESULT 11

ABL12144/c  
ID ABL12144 standard; cDNA; 5546 BP.

XX ABL12144;

AC ABL12144;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 30914.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EM;

XX WPI; 2001-658860/75.

XX P-PSDB; ABB68041.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.

XX Claim 1; SEQ ID NO 30914; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins (ABE57737-  
CC ABE72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 5546 BP; 1398 A; 1240 C; 1213 G; 1695 T; 0 U; 0 Other;

Query Match 60.6%; Score 20.6; DB 4; Length 5546;

Best Local Similarity 85.2%; Pred. No. 45;

Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 7 TACGGTTCAATTCCTCATCAGTCACCC 33

Db 2511 TCCGAGTTCAAATCCACCATCACC 2485

RESULT 12

ABZ26080/c

ID ABZ26080 standard; DNA; 40116 BP.

XX ABZ26080;

XX 24-MAR-2003 (first entry)

XX Mouse DNaseX encoding genomic DNA SEQ ID NO 7.

XX Mouse; DNase X; apoptosis; systemic lupus erythematosus; cystic fibrosis;  
KW acquired immune deficiency syndrome; AIDS; cancer; prostatic atrophy;  
KW transgenic; dermatological; immunosuppressive; anti-inflammatory;  
KW anti-HIV; cytostatic; gene; ds.



XX Thalassiosira weissflogii plastid tRNA, SEQ ID NO:168.  
XX sra gene; tRNA; bacterium; chloroplast; diatom; detection; detection;  
KW identification; quantification; characterisation; nucleic acid array;  
KW DNA chip; drug design; treatment monitoring; contamination; ss.  
XX Thalassiosira weissflogii.  
XX WO200070086-A1.  
XX 23-NOV-2000.  
XX 15-MAY-2000; 2000WO-IB000066.  
XX 14-MAY-1999; 99WO-IB000043.  
XX (IRBI-) ENTERPRISE IRELAND T/A BIORESEARCH IRELA.  
XX (DNA-) UNIV NAI IRELAND GALWAY.  
XX Barry TG, Smith TJ;  
XX WPI; 2001-025025/03.  
XX Use of sra gene, tRNA, or fragments of them, as target regions in probe  
PT assays for detection of prokaryotic or eukaryotic organisms, and for  
PT determination of species.  
XX Claim 27; Page 83; 22ipp; English.  
XX The invention relates to the sra gene or tRNA, an RNA transcript of the  
CC sra gene, or fragments thereof as target regions in a nucleic acid probe  
CC assay for the detection and identification of prokaryotic and/or  
CC eukaryotic organisms. The invention also relates to 38 novel sra  
CC sequences and their tRNA transcripts (claimed), and to pan-bacterial,  
CC genus- and species-specific sra gene/tRNA-directed probes and PCR  
CC primers (claimed). tRNA is a stable, high copy number RNA which is found  
CC in all bacteria and is also found in chloroplasts and diatoms. It has a  
CC dual function both as a tRNA and as an mRNA and is involved in rescuing  
CC truncated mRNAs which have lost stop codons. Sra genes and tRNA  
CC sequences can be used as target regions in nucleic acid probe assays for  
CC the detection, identification, or quantification of a prokaryotic or  
CC eukaryotic organism. cDNA transcripts of tRNA molecules may also be used  
CC as probes for in vitro or in situ nucleic acid hybridisation assays. A  
CC fragment of the sra gene or a tRNA molecule corresponding to a region  
CC of high homology from the 5' end or the 3' end of the DNA molecule can be  
CC used as a universal target region in a nucleic acid probe assay, while a  
CC fragment of the sra gene or a tRNA molecule corresponding to a region  
CC of low homology can be used as a target region to distinguish between  
CC species and as a target region for the generation of genus-specific  
CC probes. These regions may also be used as the basis for amplification  
CC primer design. The target regions may be used as the basis of an assay  
CC for distinguishing between living and dead prokaryotic or eukaryotic  
CC organisms, and in a multiple probe format for broad scale detection  
CC and/or identification of prokaryotic or eukaryotic organisms. An sra  
CC gene probe or a tRNA transcript probe can be linked to a microarray gene  
CC chip system for the broad scale high throughput detection and  
CC identification of prokaryotic or eukaryotic organisms. A fragment of the  
CC sra gene or the tRNA transcript can be used in an assay to obtain a DNA  
CC profile of a prokaryotic organism and distinguish between strains of the  
CC same species. The sra gene, the tRNA transcript, DNA complementary to  
CC an sra gene or tRNA, or a fragment thereof can be used to design an  
CC agent directed against infectious prokaryotic or eukaryotic organisms for  
CC therapeutic purposes, and target regions may be used to monitor the  
CC efficacy of drug therapies against infectious agents. Target regions may  
CC also be used to monitor the viability and level of probiotic organisms in  
CC the gastrointestinal tract. The methods and nucleic acids and  
CC compositions of the invention have applications in medicine, and also in  
CC industry (e.g., for assessing bacterial contamination of a foodstuff or  
CC an environmental sample). Sequences AAF15443-F15547 represent tRNAs, or  
CC fragments thereof, from a wide variety of organisms  
XX Sequence 348 BP; 124 A; 47 C; 49 G; 0 T; 128 U; 0 Other;

Query Match 57.6%; Score 19.6; DB 5; Length 348;  
Best Local Similarity 61.5%; Pred. No. 77;  
Matches 16; Conservative 6; Mismatches 4; Indels 0; Gaps 0;  
QY 8 ACGGGTTCAATTCCTCATCAGTCACCC 33  
DB 321 AUGGGUCAAUCCCAUCAGUCCAC 346  
RESULT 15  
ACH26679/C  
ID ACH26679 standard; cDNA; 438 BP.  
XX AC ACH26679;  
XX DT 13-OCT-2003 (first entry)  
XX DE Human adult ovary cDNA #5059.  
XX KW Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;  
KW genome mapping; biodiversity; genetic disorder.  
XX OS Homo sapiens.  
XX PN US2003073623-A1.  
XX PD 17-APR-2003.  
XX PF 30-JUL-2001; 2001US-00918995.  
XX PR 30-JUL-2001; 2001US-00918995.  
XX PA (DRMA/) DRMANAC R T.  
XX PA (LABA/) LABAT I.  
XX PA (STAC/) STACHE-CRAIN B.  
XX PA (DICK/) DICKSON M C.  
XX PA (JONE/) JONES L W.  
XX PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;  
XX WPI; 2003-615964/58.  
XX New polynucleotide sequences obtained from various cDNA libraries, useful  
PT as hybridization probes, as oligomers for PCR, for chromosome and gene  
PT mapping, in the recombinant production of protein, or in generating  
PT antisense DNA or RNA.  
XX Claim 1; SEQ ID NO 13891; 44pp; English.  
XX The invention relates to an isolated polynucleotide comprising any one of  
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was  
CC determined by the technique of SBH (sequencing by hybridisation). Also  
CC included is a purified polypeptide comprising a sequence corresponding to  
CC a reading frame of the novel polynucleotide. The nucleic acid sequences  
CC are useful in diagnostics as expressed sequence tags (EST) for  
CC identifying expressed genes or for physical mapping of the human genome,  
CC in forensics, in assessing biodiversity, or in identifying mutations  
CC responsible for genetic disorders and other traits. The nucleotide  
CC sequences are also useful as hybridisation probes, as oligomers for PCR,  
CC for chromosome and gene mapping, in the recombinant production of  
CC protein, or in generating antisense DNA or RNA. The purified polypeptide  
CC is useful for generating antibodies specific for it. The present sequence  
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=20030073623  
XX Sequence 438 BP; 150 A; 73 C; 81 G; 131 T; 0 U; 3 Other;  
SQ Query Match 57.6%; Score 19.6; DB 8; Length 438;  
Best Local Similarity 84.6%; Pred. No. 80;  
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 5 GGTACGGGTTCAATTCCTCAGTCA 30  
|||  
Db 313 GGTTCGGTTCATTCCTCAGTCA 288  
|||

Search completed: August 6, 2004, 09:47:34  
Job time : 23.205 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2004, 09:31:14 ; Search time 4.08664 Seconds  
(without alignments)  
4617.079 Million cell updates/sec

Title: US-09-855-340A-3

Perfect score: 34

Sequence: 1 cccggtagcggttcaattcccatcagtcacccg 34

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTCUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22.4	65.9	580073	4	US-08-545-528D-1
2	19.6	57.6	640681	4	US-09-790-988-1
3	19.2	56.5	306	4	US-09-134-000C-892
4	19.2	56.5	839	4	US-09-711-164-7
5	18.8	55.3	596	2	US-08-815-175-2
6	18.8	55.3	6305	4	US-09-221-017B-256
7	18.8	55.3	9511	1	US-07-925-695-6
8	18.8	55.3	9511	1	US-07-925-695-7
9	18.6	54.7	1538	4	US-09-644-460-28
10	18.6	54.7	2463	1	US-08-307-444A-10
11	18.6	54.7	2463	1	US-08-587-389-10
12	18.6	54.7	3466	4	US-09-880-484D-1
13	18.6	54.7	3653	4	US-09-023-655-1326
14	18.6	54.7	35100	2	US-08-770-379-17
15	18.6	54.7	35100	3	US-08-757-669A-17
16	18.6	54.7	35100	4	US-09-230-371A-17
17	18.6	54.7	4403765	3	US-09-103-840A-2
18	18.6	54.7	4411529	3	US-09-103-840A-1
19	18.4	54.1	1894	3	US-09-329-350-32
20	18.2	53.5	1281	4	US-09-489-039A-5674
21	18.2	53.5	1356	4	US-09-489-039A-359
22	18.2	53.5	2751	1	US-08-153-848-23
23	18.2	53.5	2751	3	US-09-299-843A-23
24	18.2	53.5	2751	4	US-09-088-337B-23
25	18.2	53.5	2751	5	PCT-US93-11153-23
26	18.2	53.5	5039	4	US-09-386-816C-1
27	18	52.9	472	4	US-09-548-938A-9

#### ALIGNMENTS

##### RESULT 1

US-08-545-528D-1

; Sequence 1, Application US/08545528D

; Patent No. 6537773

; GENERAL INFORMATION:

; APPLICANT: Fraser et al.

; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment

; Patent No. 6537773

; TITLE OF INVENTION: Thereof, and Uses Thereof

; FILE REFERENCE: EPI93P1

; CURRENT APPLICATION NUMBER: US/08/545,528D

; CURRENT FILING DATE: 1995-10-19

; PRIOR APPLICATION NUMBER: US 08/488,018

; PRIOR FILING DATE: 1995-06-07

; PRIOR APPLICATION NUMBER: US 08/473,545

; PRIOR FILING DATE: 1995-06-07

; NUMBER OF SEQ ID NOS: 1

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 1

; LENGTH: 580073

; TYPE: DNA

; ORGANISM: Mycoplasma genitalium

US-08-545-528D-1

Query Match 65.9%; Score 22.4; DB 4; Length 580073;

Best Local Similarity 95.8%; Pred. No. 2.5;

Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 GGCTTCAATTCCTCATCAGTCACCC 33

Db 445128 GGCTTCAATTCCTCATCAGTCACCC 445151

##### RESULT 2

US-09-790-988-1/c

; Sequence 1, Application US/09790988

; Patent No. 6632935

; GENERAL INFORMATION:

; APPLICANT: SHIGENOBU, SHUJI

; APPLICANT: WATANABE, HIDEMI

; APPLICANT: HATTORI, MASAHIRA

; APPLICANT: SAKAKI, YOSHIYUKI

; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS

; FILE REFERENCE: 081356/0159

; CURRENT APPLICATION NUMBER: US/09/790,988

; CURRENT FILING DATE: 2001-02-23

; PRIOR APPLICATION NUMBER: JP2000-107160

; PRIOR FILING DATE: 2000-04-07

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: Patentin Ver. 2.1

Sequence 1, Appli  
Sequence 1885, Ap  
Sequence 4617, Ap  
Sequence 4623, Ap  
Sequence 4636, Ap  
Sequence 5524, Ap  
Sequence 6553, Ap  
Sequence 6556, Ap  
Sequence 802, App  
Sequence 7369, Ap  
Sequence 1517, Ap  
Sequence 1586, Ap  
Sequence 4436, Ap  
Sequence 4620, Ap  
Sequence 4627, Ap  
Sequence 5474, Ap  
Sequence 6461, Ap  
Sequence 3, Appli

; SEQ ID NO 1  
; LENGTH: 640681  
; TYPE: DNA  
; ORGANISM: Buchnera sp.  
US-09-790-988-1

Query Match 57.6%; Score 19.6; DB 4; Length 640681;  
Best Local Similarity 84.6%; Pred. No. 51;  
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 8 ACGGTTCAATCCCATCAGTCACCC 33  
DB 625858 ATGGGTTCAAGTCCCATGACCC 625833

## RESULT 3

US-09-134-000C-892/c  
; Sequence 892, Application US/09134000C  
; Patent No. 6617156  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 032796-032  
; CURRENT APPLICATION NUMBER: US/09/134,000C  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/055,778  
; PRIOR FILING DATE: 1997-08-15  
; NUMBER OF SEQ ID NOS: 6812  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 892  
; LENGTH: 306  
; TYPE: DNA  
; ORGANISM: Enterococcus faecalis  
US-09-134-000C-892

Query Match 56.5%; Score 19.2; DB 4; Length 306;  
Best Local Similarity 87.5%; Pred. No. 9.8;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 10 GGGTTCATTCCTCCATCAGTCACCC 33  
DB 251 GGGTTCGATTCCTCCATCGGTCGCC 228

## RESULT 4

US-09-711-164-7  
; Sequence 7, Application US/09711164  
; Patent No. 6589738  
; GENERAL INFORMATION:  
; APPLICANT: Forsyth, R. Allyn  
; APPLICANT: Ohlsen, Karl  
; APPLICANT: Zyskind, Judith  
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERAPY  
; FILE REFERENCE: ELTRA.008A  
; CURRENT APPLICATION NUMBER: US/09/711,164  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: US 60/164415  
; PRIOR FILING DATE: 1999-11-9  
; NUMBER OF SEQ ID NOS: 469  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7  
; LENGTH: 839  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)...(839)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-711-164-7

Query Match 56.5%; Score 19.2; DB 4; Length 839;  
Best Local Similarity 72.7%; Pred. No. 13;

Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 2 CCGGTACGGTTCAATCCCATCAGTCACCC 34  
DB 640 CNGGTCAGGGTTCAATACCATCGGTAATCG 672

## RESULT 5

US-08-815-175-2  
; Sequence 2, Application US/08815175  
; Patent No. 5856139  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Goli, Surya K.  
; TITLE OF INVENTION: NOVEL PROLINE-RICH ACIDIC PROTEIN  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: US  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA: US/08/815,175  
; APPLICATION NUMBER: US/08/815,175  
; FILING DATE: Filed Herewith  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0225 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 596 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: PANTUT02  
; CLONE: 2235738  
US-08-815-175-2

Query Match 55.3%; Score 18.8; DB 2; Length 596;  
Best Local Similarity 69.7%; Pred. No. 18;  
Matches 23; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 CCGGTACGGTTCAATCCCATCAGTCACCC 33  
DB 497 CCCAGTAGGGTTCCAGGGCCATNAGTGNCCC 529

## RESULT 6

US-09-221-017B-256/c  
; Sequence 256, Application US/09221017B  
; Patent No. 6444799  
; GENERAL INFORMATION:  
; APPLICANT: Ross, Bruce C.  
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
; NUMBER OF SEQUENCES: 1120  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER



```
/ FILING DATE: 09-AUG-1991
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: JP 360441/91
/ FILING DATE: 05-DEC-1991
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Weillacher, Robert G.
/ REGISTRATION NUMBER: 20,531
/ REFERENCE/DOCKET NUMBER: 06/87-48009
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (202) 659-2811
/ TELEFAX: (202) 659-1462
/ TELEX: WUI 64470
/ INFORMATION FOR SEQ ID NO: 7:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 9511 base pairs
/ TYPE: NUCLEIC ACID
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ US-07-925-695-7

Query Match 55.3%; Score 18.8; DB 1; Length 9511;
Best Local Similarity 76.7%; Pred. No. 39;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 CCGGTACGGGTTCATTCCTCCATCAGTCAC 31
Db 4407 CCGGTACGGTGACACACTCCACAGTAAC 4436

RESULT 9
US-09-644-460-28/c
/ Sequence 28, Application US/09644460
/ Patent No. 6657053
/ GENERAL INFORMATION:
/ APPLICANT: Fisher, Paul B.
/ TITLE OF INVENTION: Reciprocal Subtraction Differential
/ TITLE OF INVENTION: Display
/ FILE REFERENCE: 34587-C-PCT-USA
/ CURRENT APPLICATION NUMBER: US/09/644,460
/ CURRENT FILING DATE: 2000-08-23
/ PRIOR APPLICATION NUMBER: PCT/US99/04323
/ PRIOR FILING DATE: 1999-02-26
/ PRIOR APPLICATION NUMBER: US 09/197,889
/ PRIOR FILING DATE: 1998-11-23
/ PRIOR APPLICATION NUMBER: US 09/185,115
/ PRIOR FILING DATE: 1998-11-03
/ PRIOR APPLICATION NUMBER: US 09/032,684
/ PRIOR FILING DATE: 1998-02-27
/ NUMBER OF SEQ ID NOS: 42
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 28
/ LENGTH: 1538
/ TYPE: DNA
/ ORGANISM: homo sapiens
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: 652, 1523
/ OTHER INFORMATION: c, t, a or g
/ NAME/KEY: misc.feature
/ LOCATION: (1)-(1538)
/ OTHER INFORMATION: n = A,T,C or G
US-09-644-460-28

Query Match 54.7%; Score 18.6; DB 4; Length 1538;
Best Local Similarity 72.7%; Pred. No. 29;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 CCCCCGTCAGGTTCATTCCTCCATCAGTCACCC 33
Db 1240 CCGGTACGGTGACAGTCCCTCAGACGCC 1208

RESULT 10
```

```
US-08-307-444A-10
/ Sequence 10, Application US/08307444A
/ Patent No. 5516659
/ GENERAL INFORMATION:
/ APPLICANT: NII, ATSUSHI
/ APPLICANT: MORISHITA, HIDEAKI
/ APPLICANT: UEMURA, AKIO
/ APPLICANT: MOCHIDA, EI
/ TITLE OF INVENTION: ANTICOAGULANT POLYPEPTIDES
/ NUMBER OF SEQUENCES: 27
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: OLIFF & BERRIDGE
/ STREET: P.O. BOX 19928
/ CITY: ALEXANDRIA
/ STATE: VA
/ COUNTRY: USA
/ ZIP: 22320
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/307,444A
/ FILING DATE: 19-SEP-1994
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/835,436
/ FILING DATE: 26-FEB-1992
/ ATTORNEY/AGENT INFORMATION:
/ NAME: OLIFF, JAMES A.
/ REGISTRATION NUMBER: 27,075
/ REFERENCE/DOCKET NUMBER: JAO 27706
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (703) 836-8400
/ TELEFAX: (703) 836-2787
/ TELEX: 90-1799 PTO ALEX
/ INFORMATION FOR SEQ ID NO: 10:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2463 bases
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
US-08-307-444A-10

Query Match 54.7%; Score 18.6; DB 1; Length 2463;
Best Local Similarity 72.7%; Pred. No. 33;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 CCCCCGTCAGGTTCATTCCTCCATCAGTCACCC 33
Db 2338 CCGAGGAAGTGGCCCACTCAGTCAGTCACCC 2370

RESULT 11
US-08-587-389-10
/ Sequence 10, Application US/08587389
/ Patent No. 5695964
/ GENERAL INFORMATION:
/ APPLICANT: NII, ATSUSHI
/ APPLICANT: MORISHITA, HIDEAKI
/ APPLICANT: UEMURA, AKIO
/ APPLICANT: MOCHIDA, EI
/ TITLE OF INVENTION: TRUNCATED THROMBOMODULIN, RECOMBINANT
/ NUMBER OF SEQUENCES: 27
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: OLIFF & BERRIDGE
/ STREET: P.O. BOX 19928
/ CITY: ALEXANDRIA
```

STATE: VA  
COUNTRY: USA  
ZIP: 22320  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: US/08/567,389  
APPLICATION NUMBER: US/08/567,389  
FILING DATE: 17-JAN-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/307,444  
FILING DATE: 19-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: OLIFF, JAMES A.  
REGISTRATION NUMBER: 27,075  
REFERENCE/DOCKET NUMBER: JAO 27706  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6400  
TELEFAX: (703) 836-2787  
TELEX: 90-1799 PTO ALEX  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2463 bases  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-587-389-10

Query Match 54.7%; Score 18.6; DB 1; Length 2463;  
Best Local Similarity 72.7%; Pred. No. 33;  
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 CCCCGGTACGGGTTCAATTCCCATCAGTCACCC 33  
DB 2338 CCCAGGAAGTGGGCCAACTCACCTGAGTCACCC 2370

RESULT 12  
US-09-880-484D-1  
; Sequence 1, Application US/09880484D  
; Patent No. 6632791  
; GENERAL INFORMATION:  
; APPLICANT: Light, David  
; APPLICANT: Nagashima, Mariko  
; APPLICANT: Morser, Michael J.  
; TITLE OF INVENTION: Thrombomodulin Analogs for Pharmaceutical Use  
; FILE REFERENCE: 51863AUSM1  
; CURRENT APPLICATION NUMBER: US/09/880,484D  
; CURRENT FILING DATE: 2001-06-12  
; PRIOR APPLICATION NUMBER: 60/213,678  
; PRIOR FILING DATE: 2000-06-21  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 1  
; LENGTH: 3466  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (151)..(1875)  
; OTHER INFORMATION:  
; FEATURE:  
; NAME/KEY: mat peptide  
; LOCATION: (205)..()  
; OTHER INFORMATION:  
US-09-880-484D-1

Query Match 54.7%; Score 18.6; DB 4; Length 3466;  
Best Local Similarity 72.7%; Pred. No. 36;  
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 CCCCGGTACGGGTTCAATTCCCATCAGTCACCC 33  
DB 2450 CCCAGGAAGTGGGCCAACTCACCTGAGTCACCC 2482

RESULT 13  
US-09-023-655-1326  
; Sequence 1326, Application US/09023655  
; Patent No. 6607879  
; GENERAL INFORMATION:  
; APPLICANT: Cocks, Benjamin G.  
; APPLICANT: Susan G. Stuart  
; APPLICANT: Jeffrey J. Seilhamer  
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE  
; TITLE OF INVENTION: EXPRESSION  
; NUMBER OF SEQUENCES: 1508  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/023,655  
; FILING DATE: HEREWITH  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zeller, Karen J.  
; REGISTRATION NUMBER: 37,071  
; REFERENCE/DOCKET NUMBER: PA-0001 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555  
; TELEFAX: (650) 845-4166  
; INFORMATION FOR SEQ ID NO: 1326:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3653 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GENBANK  
; CLONE: g339656  
; US-09-023-655-1326

Query Match 54.7%; Score 18.6; DB 4; Length 3653;  
Best Local Similarity 72.7%; Pred. No. 37;  
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 CCCCGGTACGGGTTCAATTCCCATCAGTCACCC 33  
DB 2446 CCCAGGAAGTGGGCCAACTCACCTGAGTCACCC 2478

RESULT 14  
US-08-770-379-17/c  
; Sequence 17, Application US/08770379  
; Patent No. 5849564  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Yuan

APPLICANT: Bohenzky, Roy A.  
APPLICANT: Russo, James J.  
APPLICANT: Edelman, Isidore S.  
APPLICANT: Moore, Patrick S.  
TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED  
HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/770,379  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 52342  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35100 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-770-379-17

Query Match 54.7%; Score 18.6; DB 2; Length 35100;  
Best Local Similarity 72.7%; Pred. No. 69;  
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 1 CCCCGGTACGGGTCAATTCCCATCAGTCACCC 33  
Db 7143 CCCAGGTACGGGTGAGATACCCCTCAGCTTCCC 7111

RESULT 15  
US-08-757-669A-17/c  
Sequence 17, Application US/08/757669A  
Patent No. 6183751  
GENERAL INFORMATION:  
APPLICANT: Chang, Yuan  
APPLICANT: Bohenzky, Roy A.  
APPLICANT: Russo, James J.  
APPLICANT: Edelman, Isidore S.  
APPLICANT: Moore, Patrick S.  
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/757,669A  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 45185-F  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35100 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-757-669A-17  
Query Match 54.7%; Score 18.6; DB 3; Length 35100;  
Best Local Similarity 72.7%; Pred. No. 69;  
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 1 CCCCGGTACGGGTCAATTCCCATCAGTCACCC 33  
Db 7143 CCCAGGTACGGGTGAGATACCCCTCAGCTTCCC 7111  
Search completed: August 6, 2004, 13:42:37  
Job time : 10.0866 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2004, 09:22:34 ; Search time 131.685 Seconds  
(without alignments)  
7710.163 Million cell updates/sec

Title: US-09-855-340A-3  
Perfect score: 34  
Sequence: 1 cccggtagcgggttcattccatccatgcagtcacccg 34

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Lasting first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
C 1	22.4	65.9	689	28	AZ933974 BJ_Ba000
C 2	21.8	64.1	663	10	BF165844 60174575
C 3	21.8	64.1	913	13	BQ889122 AGENCOURT
C 4	21.8	64.1	914	10	BF540384 602050148

C	5	21.4	62.9	496	14	CD221929	CCCL1_E0
C	6	21.4	62.9	617	28	BZ253996	CH230-307
C	7	21.4	62.9	619	14	CD228161	CCCL1_60
C	8	21.2	62.4	512	28	AZ663671	IM0543K22
C	9	21.2	62.4	647	28	BH698487	BOHXM56TR
C	10	21.2	62.4	649	28	AZ663662	IM0543I24
C	11	20.8	61.2	369	28	AZ658076	IM0534H08
C	12	20.8	61.2	435	29	CG962621	MEEFX20TR
C	13	20.8	61.2	459	28	BZ755997	PUEFL38TD
C	14	20.8	61.2	481	28	BZ866579	CH240_287
C	15	20.8	61.2	503	12	BI128480	G076P43Y
C	16	20.8	61.2	607	28	BZ440539	BONNF14TF
C	17	20.8	61.2	772	10	BF166788	601775109
C	18	20.8	61.2	772	28	CG335988	OGPAW05TH
C	19	20.8	61.2	797	29	CG963642	MEEVL9TR
C	20	20.8	61.2	842	29	CG026738	PGAA19TF
C	21	20.8	61.2	846	29	CG932119	MBENG41TR
C	22	20.8	61.2	881	28	BZ405594	OGABA83TM
C	23	20.8	61.2	882	12	BI736662	603360246
C	24	20.8	61.2	896	29	CG945888	MBEJG36TR
C	25	20.8	61.2	931	29	CG953021	MBEPR56TR
C	26	20.8	61.2	1161	12	BG342311	602374217
C	27	20.6	60.6	459	28	AZ407486	IM0178A20
C	28	20.6	60.6	499	29	TA134G07Q	AL466345 T. brucei
C	29	20.6	60.6	589	28	AQ272293	nbxb0027D
C	30	20.6	60.6	697	28	BH105996	BH105996 RPCI-24-3
C	31	20.4	60.0	239	9	AV345146	AV345146 AV345146
C	32	20.4	60.0	274	9	AV338955	AV338955 AV338955
C	33	20.4	60.0	478	10	BE554144	BE554144 UT-M-AN1-
C	34	20.4	60.0	527	28	BH290990	CH230-114
C	35	20.4	60.0	640	10	BE634824	uv85604.X
C	36	20.4	60.0	654	10	BB484517	BB484517 BB484517
C	37	20.4	60.0	669	14	CB841703	CB841703 M15E-2276
C	38	20.4	60.0	779	13	BU562772	AGENCOURT
C	39	20.4	60.0	856	10	BB036725	BB036725 B3036725
C	40	20.4	60.0	986	9	AV327597	AV327597 AV327597
C	41	20.4	60.0	1890	11	AK031860	AK031860 Mus muscu
C	42	20.4	60.0	2893	11	AK019872	AK019872 Mus muscu
C	43	20.4	60.0	3747	11	AK037906	AK037906 Mus muscu
C	44	20.2	59.4	247	12	BI003984	BI003984 RC4-HN004
C	45	20.2	59.4	338	9	AA343504	AA343504 EST49283

#### ALIGNMENTS

RESULT 1  
AZ933974/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
COMMENT

AZ933974  
BJ\_Ba0001H19r B. japonicum BAC library  
genomic, genomic survey sequence.  
GSS.  
GI:13776034  
Bradyrhizobium japonicum  
Bradyrhizobium japonicum  
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
Bradyrhizobiaceae; Bradyrhizobium.  
1 (bases 1 to 689)  
Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A.,  
Goicoechea,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.  
A marker-dense, sequence-ready map of the Bradyrhizobium japonicum  
Genome Res. 11 (8), 1434-1440 (2001)  
21376150  
11483585  
Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu



```

REFERENCE
AUTHORS NIH-MGC http://mgs.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLN9515 row: 0 column: 12
High quality sequence stop: 601.
FEATURES
source
Location/Qualifiers
1..914
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4189667"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP SG2"
/notes="Organ: salivary gland; Vector: pCMV-SPORT6; Site 1:
NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life
Technologies. Note: this is a NCI CGAP Library."
ORIGIN
Query Match 64.1%; Score 21.8; DB 10; Length 914;
Best Local Similarity 78.8%; Pred. No. 2.4e+02;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 1 CCGCGTAGGGTTCAATTCCCATCAGTCACCC 33
|||||
DB 602 CCGCGTAGCCCTACAAATTCCTCAGTCCTCC 634

RESULT 5
CD221929/c
LOCUS CD221929 496 bp mRNA linear EST 21-MAY-2003
DEFINITION CCC1_1_E04_b1_A007 Callus culture/cell suspension Sorghum bicolor
cDNA clone CCC1_1_E04_A007 3', mRNA sequence.
ACCESSION CD221929
VERSION CD221929.1 GI:30965363
KEYWORDS EST.
SOURCE Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE
1 (bases 1 to 496)
Cordonnier-Pratt, M.-M., Wentzel, V., Suzuki, Y., Sugano, S.,
Klein, R.R., Liang, C., Sun, F., Sullivan, R., Shah, M., Rathore, K.,
Eastman, A. and Pratt, L.H.
An EST database from Sorghum: callus culture and cell suspension
Unpublished (2003)
Other ESTs: CCC1_1_E04_G1_A007
Contact: Cordonnier-Pratt, M.
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in
the Human Genome Center, University of Tokyo Institute of Medical
Science; plant material and RNA prepared at Texas A & M University;
sequencing done in the Laboratory for Genomics and Bioinformatics,
University of Georgia. Sequence ends have been trimmed to exclude
vector and regions below Phred quality 16. Three-prime sequences
are presented as their reverse complement and have been trimmed to

```

```

exclude polyA.
Seq primer: Sug3 (CGACCTGCAGCTCGAGCACA)
POLYA=Yes.
FEATURES
source
Location/Qualifiers
1..496
/organism="Sorghum bicolor"
/mol_type="mRNA"
/cultivar="RTX430"
/db_xref="taxon:4558"
/clone="CCC1_1_E04_A007"
/lab_host="DH10B-T1 phage-resistant E. coli"
/clone_lib="Gallus culture/cell suspension"
/notes="Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; The
library was prepared from a mixture of polyA+ RNA from
callus culture tissue and cells in suspension culture.
Double-stranded cDNA was cloned unidirectionally into
different DraIII sites of the pME18S-FL3 vector (5-prime
DraIII site is CACTGTGTG, 3-prime DraIII site is
CAACCATGTG). XhoI excises the cDNA insert."
ORIGIN
Query Match 62.9%; Score 21.4; DB 14; Length 496;
Best Local Similarity 80.6%; Pred. No. 3.2e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 3 CCGGTAGGGTTCAATTCCTCAGTCACCC 33
|||||
DB 255 CCGCACAGGTTCGATTCCTCAGTCCTCACC 225

RESULT 6
BZ253996
LOCUS BZ253996 617 bp DNA linear GSS 15-OCT-2002
DEFINITION CH230-307C21-TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
CH230-307C21, genomic survey sequence.
ACCESSION BZ253996
VERSION BZ253996.1 GI:23957934
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 617)
Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K.,
Shvartsbeyn, A., Gebregeorgis, E., Overton, L., Russell, D., Chen, D.,
Riggs, F., de Jong, P. and Fraser, C.M.
Rat BAC End Sequences from Library CHORI-230 MboI segment
Unpublished (1999)
Other GSSs: CH230-307C21.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/orering\_information.htm). BAC end
page: http://www.tigr.org/tadb/bac\_ends/rat/bac\_end\_intro.html
Plate: 307 row: C column: 21
Seq primer: T7
Class: BAC ends.
FEATURES
source
Location/Qualifiers
1..617
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SsNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-307C21"

```



Best Local Similarity 88.5%; Pred. No. 3.9e+02; Mismatches 3; Indels 0; Gaps 0;  
 Matches 23; Conservative 0;

QY 6 GTACGGTTCAATCCCATCAGTCAC 31  
 |||||  
 Db 299 GAACGTGTTCAATCCCATCAGTCAC 324

RESULT 9  
 BH698487  
 LOCUS BOHXM56TR BO\_2.3\_KB Brassica oleracea genomic clone BOHXM56, linear GSS 20-FEB-2002  
 DEFINITION genomic survey sequence.

ACCESSION BH698487  
 VERSION  
 KEYWORDS  
 SOURCE GSS.

ORGANISM Brassica oleracea

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 647)

Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.

Whole genome shotgun sequencing of Brassica oleracea

Unpublished (2001)

Other GSSs: BOHXM56TF

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

LOCATION/Qualifiers

1..647

/organism="Brassica oleracea"

/mol\_type="genomic DNA"

/strain="TO1000DH3"

/db\_xref="taxon:3712"

/clone="BOHXM56"

/clone\_lib="BO\_2.3\_KB"

/note="Vector: pHO51; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHO51 using BstXI linkers"

ORIGIN

Query Match 62.4%; Score 21.2; DB 28; Length 647;

Best Local Similarity 88.5%; Pred. No. 4e+02; Mismatches 3; Indels 0; Gaps 0;

Matches 23; Conservative 0;

QY 9 CGGGTTCAATCCCATCAGTCACCG 34

|||  
 Db 618 CGGGTTCAATCCCATCAGTCACCG 643

RESULT 10

AZ663662

LOCUS 1M0543124F Mouse 10kb plasmid UUGC1M library Mus musculus genomic

DEFINITION clone UUGC1M0543124 F, genomic survey sequence.

ACCESSION AZ663662

VERSION AZ663662.1 GI:11800808

KEYWORDS GSS.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 649)

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, I.,

Reilly, M., Rose, R., Rose, R., Stokes, R., Tingey, A., von

Niederhausern, A. and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)

COMMENT Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0543 row: I column: 24

Seq primer: CGTGTAAACGACGCGCAGT

Class: plasmid ends

High quality sequence stop: 649.

LOCATION/Qualifiers

1..649

/organism="Mus musculus"

/mol\_type="genomic DNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC1M0543124"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"

/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The

adaptor DNA was purified and size-selected for a 9.5 to 10.5 Kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gi|4732114|gb|AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and

purified. The sheared, adaptor mouse DNA was annealed to

adaptor vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

ORIGIN

Query Match 62.4%; Score 21.2; DB 28; Length 649;

Best Local Similarity 88.5%; Pred. No. 4e+02; Mismatches 3; Indels 0; Gaps 0;

Matches 23; Conservative 0;

QY 6 GTACGGTTCAATCCCATCAGTCAC 31

|||  
 Db 300 GAACGTGTTCAATCCCATCAGTCAC 325

RESULT 11

AZ658076

LOCUS 1M0534H08R Mouse 10kb plasmid UUGC1M library Mus musculus genomic

DEFINITION clone UUGC1M0534H08 R, genomic survey sequence.

ACCESSION AZ658076

VERSION AZ658076.1 GI:11795222

KEYWORDS GSS.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 369)

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, I.,

Reilly, M., Rose, R., Rose, R., Stokes, R., Tingey, A., von

Niederhausern, A. and Wright, D., Weiss, R.

**TITLE** Mouse whole genome scaffolding with paired end reads from 10kb  
**JOURNAL** Unpublished (2000)  
**COMMENT** Contact: Robert B. Weiss  
 University of Utah Genome Center  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: rdunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0534 row: H column: 08  
 Seq primer: CACACAGGAAACAGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 369.

**FEATURES**  
 source  
 1. 369  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0534H08"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, TI-resistant, F-"  
 /notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (G14732114|GB|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

**ORIGIN**  
 Query Match 61.2%; Score 20.8; DB 28; Length 369;  
 Best Local Similarity 91.7%; Pred. No. 5.4e+02;  
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

**QY** 10 GGTTCAATTCCTCAGTCAGTCACCC 33  
 |||||  
**Db** 212 GGTTCAATTCCTCAGTCAGTCACAC 235  
 |||||

**RESULT 12**  
**CG962621/c**  
**LOCUS** MBEFX20TR mch2 Medicago truncatula genomic clone 47D16, genomic survey sequence.  
**DEFINITION** CG962621  
**ACCESSION** CG962621  
**VERSION** CG962621.1 GI:39885464  
**KEYWORDS** GSS.  
**SOURCE** Medicago truncatula (barrel medic)  
**ORGANISM** Medicago truncatula  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.  
 1 (bases 1 to 425)  
 Town, C.D., Shetty, J., Koo, H. and Feldblyum, T.F.  
**AUTHORS** Sequencing of BAC ends from Medicago truncatula  
**TITLE**

**Query Match** 61.2%; Score 20.8; DB 28; Length 459;  
**Best Local Similarity** 78.1%; Pred. No. 5.6e+02;  
**Matches** 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

**QY** 2 CCCTGACGGTTCATTCATCCATCAGTCACCC 33  
 |||||  
**Db** 158 CCATGCCGCGGTTCATTCATCCGTCATTCGCC 127  
 |||||

**RESULT 13**  
**BZ755997**  
**LOCUS** PUBFL38TD ZM\_0.6\_1.0 KB Zea mays genomic clone ZMMBTA281G04, genomic survey sequence.  
**DEFINITION** BZ755997  
**ACCESSION** BZ755997.1 GI:28908378  
**VERSION** BZ755997  
**KEYWORDS** GSS.  
**SOURCE** Zea mays  
**ORGANISM** Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 459)  
 Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and Bennetzen, J.  
**AUTHORS** Maize Genomics Consortium  
 Unpublished (2003)  
**TITLE** Zea mays  
**JOURNAL** Contact: Cathy Whitelaw  
**COMMENT** TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843  
 Fax: 301-838-0208  
 Email: whitelaw@tigr.org  
 Seq primer: Tg  
 Class: sheared ends.

**FEATURES**  
 source  
 1. 459  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZMMBTA281G04"  
 /notes="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high COT selected genomic DNA library"

**ORIGIN**  
 Query Match 61.2%; Score 20.8; DB 28; Length 459;  
 Best Local Similarity 78.1%; Pred. No. 5.6e+02;  
 Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

**JOURNAL** Unpublished (2003)  
**COMMENT** Other GSSs: MBEFX20TF  
 Contact: Chris Town  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel: 301-838-3223  
 Fax: 301-838-0208  
 Email: cdtown@tigr.org  
 Seq primer: CAGGAAACAGCTATGACC  
 Class: BAC ends.

**FEATURES**  
 Location/Qualifiers  
 1. 425  
 /organism="Medicago truncatula"  
 /mol\_type="genomic DNA"  
 /cultivar="genotype A17"  
 /db\_xref="taxon:3880"  
 /clone="47D16"  
 /clone\_lib="mth2"  
 /note="Vector: pBelOBAC11; Site 1: HindIII; Site 2: HindIII; Cook, D.R. and Kim, D.J., unpublished"

**ORIGIN**  
 Query Match 61.2%; Score 20.8; DB 29; Length 425;  
 Best Local Similarity 78.1%; Pred. No. 5.5e+02;  
 Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

**QY** 2 CCCTGACGGTTCATTCATCCATCAGTCACCC 33  
 |||||  
**Db** 158 CCATGCCGCGGTTCATTCATCCGTCATTCGCC 127  
 |||||

**RESULT 13**  
**BZ755997**  
**LOCUS** PUBFL38TD ZM\_0.6\_1.0 KB Zea mays genomic clone ZMMBTA281G04, genomic survey sequence.  
**DEFINITION** BZ755997  
**ACCESSION** BZ755997.1 GI:28908378  
**VERSION** BZ755997  
**KEYWORDS** GSS.  
**SOURCE** Zea mays  
**ORGANISM** Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 459)  
 Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and Bennetzen, J.  
**AUTHORS** Maize Genomics Consortium  
 Unpublished (2003)  
**TITLE** Zea mays  
**JOURNAL** Contact: Cathy Whitelaw  
**COMMENT** TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843  
 Fax: 301-838-0208  
 Email: whitelaw@tigr.org  
 Seq primer: Tg  
 Class: sheared ends.

**FEATURES**  
 Location/Qualifiers  
 1. 459  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZMMBTA281G04"  
 /notes="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high COT selected genomic DNA library"

**ORIGIN**  
 Query Match 61.2%; Score 20.8; DB 28; Length 459;  
 Best Local Similarity 78.1%; Pred. No. 5.6e+02;  
 Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;



**This Page Blank (uspto)**